A Unified Treatment of Convexity of Relative Entropy and Related Trace Functions, with Conditions for Equality

Anna Jenčová *
Mathematical Institute, Slovak Academy of Sciences,
Štefánikova 49, 814 73 Bratislava, Slovakia
jenca@mat.savba.sk

Mary Beth Ruskai †
Department of Mathematics, Tufts University, Medford, MA 02155, USA
marybeth.ruskai@tufts.edu

July 30, 2010

Abstract

We introduce a generalization of relative entropy derived from the Wigner-Yanase-Dyson entropy and give a simple, self-contained proof that it is convex. Moreover, special cases yield the joint convexity of relative entropy, and for $\mathrm{Tr}\, K^* A^p K B^{1-p}$ Lieb's joint concavity in (A, B) for $0 < p < 1$ and Ando's joint convexity for $1 < p \leq 2$. This approach allows us to obtain conditions for equality in these cases, as well as conditions for equality in a number of inequalities which follow from them. These include the monotonicity under partial traces, and some Minkowski type matrix inequalities proved by Lieb and Carlen for $\mathrm{Tr}_1 (\mathrm{Tr}_2\, A_{12}^p)^{1/p}$. In all cases the equality conditions are independent of p; for extensions to three spaces they are identical to the conditions for equality in the strong subadditivity of relative entropy.

*Supported by the grants VEGA 2/0032/09 and APVV- 0071-06
†Partially supported by National Science Foundation under Grant DMS-0604900.

1 Introduction

1.1 Background

For matrices $A_{12} > 0$ acting on a tensor product of two Hilbert spaces, Carlen and Lieb [7, 8] considered the trace function $\left[\mathrm{Tr}_1(\mathrm{Tr}_2 A_{12}^p)^{q/p}\right]^{1/q}$ and proved that it is concave when $0 \le p \le q \le 1$ and convex when $1 \le q$ and $1 \le p \le 2$. They showed that this implies that these functions and the norms they generate satisfy Minkowski type inequalities, including a natural generalization to matrices A_{123} acting on a tensor product of three Hilbert spaces. They also raised the question of the conditions for equality in their inequalities. When $q = 1$, we show that this can be treated using methods developed to treat equality in the strong subadditivity of quantum entropy. Moreover, we obtain conditions for equality in a large class of related convexity inequalities, show that they are independent of p in the range $0 < p < 2$, and show that for inequalities involving A_{123} they are identical to the equality conditions for strong subadditivity (SSA) of quantum entropy give in [13]

These equality conditions are non-trivial and have found many applications in quantum information theory. For example, they play an important role in some recent "no broadcasting" results; see [18] and references therein. They also plays a key role in Devetak and Yard's [9] "quantum state redistribution" protocol which gives an operational interpretation to the quantum conditional mutual information.

Our approach to proving joint convexity of relative entropy is motivated by Araki's relative modular operator [5], introduced to generalize relative entropy to more general situations including type III von Neumann algebras. It was subsequently used by Narnhofer and Thirring [28] to give a new proof of SSA. The argument given here is similar to that in [17, 30, 36]; however, the unified treatment for $0 < p < 2$ leading to equality conditions, is new. Moreover, a dual treatment can be given for $-1 < p < 1$ allowing extension to the full range $(-1, 2)$.

Wigner and Yanase [41, 42] introduced the notion of skew information of a density matrix γ with respect to a self-adjoint observable K,

$$-\tfrac{1}{2}[K, \gamma^p]\,[K, \gamma^{1-p}] \tag{1}$$

for $p = \frac{1}{2}$ and Dyson suggested extending this to $p \in (0, 1)$. Wigner and Yanase [42] proved that (1) is convex in γ for $p = \frac{1}{2}$ and, in his seminal paper [19] on convex trace functions, Lieb proved joint concavity for $p \in (0, 1)$ for the more general function

$$(A, B) \mapsto \mathrm{Tr}\, K^* A^p K B^{1-p} \tag{2}$$

for K fixed and $A, B > 0$ positive semi-definite. This implies convexity of (1) and was a key step in the original proof [22] of the strong subadditivity (SSA) inequality

of quantum entropy. Moreover, it leads to a proof of joint convexity of relative entropy[1] as well. It is less well known that Ando [3, 4] gave another proof which also showed that for $1 \leq p \leq 2$, the function (2) is jointly convex in A, B. The case $p = 2$ was considered earlier by Lieb and Ruskai [23]. We modify what one might describe as Lieb's extension of the Wigner-Yanase-Dyson (WYD) entropy to a type of relative entropy in a way that allows a unified treatment of the convexity and concavity of $\operatorname{Tr} K^* A^p K B^{1-p}$ in the range $p \in (0, 2]$ and includes the usual relative entropy as a special case. Our modification retains a linear term, even for $A \neq B$. Although this might seem unnecessary for convexity and concavity questions, it is crucial to a unified treatment.

Lieb also considered $\operatorname{Tr} K^* A^p K B^q$ with $p, q > 0$ and $0 \leq p + q \leq 1$ and Ando considered $1 < q \leq p \leq 2$. In Section 2.2, we extend our results to this situation. However, we also show that for $q \neq 1-p$, equality holds only under trivial conditions. Therefore, we concentrate on the case $q = 1 - p$.

Next, we introduce our notation and conventions. In Section 2, we first describe our generalization of relative entropy and prove its convexity; then consider the extension to $q \neq 1-p$ mentioned above; and finally prove monotonicity under partial traces including a generalization of strong subadditivity to $p \neq 1$. In Section 3, we consider several formulations of equality conditions. In Section 4, we show how to use these results to obtain equality conditions in the results of Lieb and Carlen [7, 8]. For completeness, we include an appendix which contains the proof of a basic convexity result from [36] that is key to our results.

1.2 Notation and conventions

We introduce two linear maps on the space M_d of $d \times d$ matrices. Left multiplication by A is denoted L_A and defined as $L_A(X) = AX$; right multiplication by B is denoted R_B and defined as $R_B(X) = XR$. These maps are associated with the relative modular operator $\Delta_{AB} = L_A R_B^{-1}$ introduced by Araki in a far more general context. They have the following properties:

a) The operators L_A and R_B commute since

$$L_A[R_B(X)] = AXB = R_B[L_A(X)] \tag{3}$$

even when A and B do not commute.

[1] In [22] only concavity of the conditional entropy was proved explicitly, but the same argument [35, Section V.B] yields joint convexity of the relative entropy. Independently, Lindblad [25] observed that this follows directly from (2) by differentiating at $p = 1$.

b) L_A and R_A are invertible if and only if A is non-singular, in which case $L_A^{-1} = L_{A^{-1}}$ and $R_A^{-1} = R_{A^{-1}}$.

c) When A is self-adjoint, L_A and R_A are both self-adjoint with respect to the Hilbert Schmidt inner product $\langle A, B\rangle = \operatorname{Tr} A^* B$.

d) When $A \geq 0$, the operators L_A and R_A are positive semi-definite, i.e.,

$$\begin{aligned} \operatorname{Tr} X^* L_A(X) &= \operatorname{Tr} X^* A X \geq 0 \qquad \text{and} \\ \operatorname{Tr} X^* R_A(X) &= \operatorname{Tr} X^* X A = \operatorname{Tr} X A X^* \geq 0. \end{aligned}$$

e) When $A > 0$, then $(L_A)^p = L_{A^p}$ and $(R_A)^p = R_{A^p}$ for all $p \geq 0$. If A is also non-singular, this extends to all $p \in \mathbf{R}$. More generally $f(L_A) = L_{f(A)}$ for $f : (0, \infty \mapsto)\mathbf{R}$.

To see why (e) holds, it suffices to observe that $A > 0$ implies L_A and R_A are linear operators for which $f(A)$ can be defined by the spectral theorem for any function f with domain in $(0, \infty)$. It is easy to verify that $A|\phi_j\rangle = \alpha_j|\phi_j\rangle$ implies $L_A|\phi_j\rangle\langle\phi_k| = \alpha_j|\phi_j\rangle\langle\phi_k|$ for $k = 1 \ldots d$ so that the spectral decomposition of A induces one on L_A with degeneracy d and $f(L_A)|\phi_j\rangle\langle\phi_k| = f(\alpha_k)|\phi_j\rangle\langle\phi_k|$. For R_B a similar argument goes through starting with left eigenvectors of B i.e., $\langle\phi_j|B = \beta_j\langle\phi_j|$.

If a function is homogeneous of degree 1, then convexity is equivalent to subadditivity. Thus, if $F(\lambda A) = \lambda F(A)$, then F is convex if and only if $F(A) \leq \sum_j F(A_j)$ with $A = \sum_j A_j$. We will use this equivalence without further ado.

We will encounter expressions involving commuting positive semi-definite matrices A, D with $\ker D \subseteq \ker A$. We will simply write AD^{-1} for

$$\lim_{\epsilon \to 0} \sqrt{A}(D + \epsilon I)^{-1}\sqrt{A} = \begin{cases} A\big|_{(\ker A)^\perp} \big(D\big|_{(\ker A)^\perp}\big)^{-1} & \text{on } (\ker A)^\perp \\ 0 & \text{on } \ker A \end{cases} \tag{4}$$

For B positive semi-definite, we denote the projection onto $(\ker B)^\perp$ by $P_{(\ker B)^\perp}$.

2 WYD entropy revisited and extended

2.1 Generalization of relative entropy

We now introduce the family of functions

$$g_p(x) = \begin{cases} \frac{1}{p(1-p)}(x - x^p) & p \neq 1 \\ x \log x & p = 1 \end{cases}. \tag{5}$$

which are well-defined for $x > 0$ and $p \neq 0$. For our purposes, it would suffice to consider $p \in [\frac{1}{2}, 2]$. For A, B strictly positive we define

$$\begin{aligned} J_p(K,A,B) &\equiv \operatorname{Tr} \sqrt{B} K^* \, g_p\big(L_A R_B^{-1}\big)(K\sqrt{B}) && (6) \\ &= \begin{cases} \frac{1}{p(1-p)}\big(\operatorname{Tr} K^*AK - \operatorname{Tr} K^*A^pKB^{1-p}\big) & p \in (0,1)\cup(1,2) \\ \operatorname{Tr} KK^*A\log A - \operatorname{Tr} K^*AK \log B) & p = 1 \\ -\frac{1}{2}\big(\operatorname{Tr} K^*AK - \operatorname{Tr} AKB^{-1}K^*A & p = 2 \end{cases} && (7) \end{aligned}$$

When $p = 1$ and $K = I$, (6) reduces to the usual relative entropy, i.e.,

$$J_1(I,A,B) = H(A,B) = \operatorname{Tr} A(\log A - \log B) \tag{8}$$

For $p \neq 1$, the function $J_p(K,A,B)$ differs from that considered by Lieb [19] and Ando [3, 4] by the seemingly irrelevant linear term $\operatorname{Tr} K^*AK$ and the factor $\frac{1}{p(1-p)}$. However, this minor difference allows us to give a unified treatment of $p \in (0,2]$ because of the extension by continuity to $p = 1$ and the sign change there.

One might expect to associate the exchange $A \leftrightarrow B$ with the symmetry $p \leftrightarrow (1-p)$ around $p = \frac{1}{2}$. However, this is problematic at $p = 1$. Therefore, we use instead the observation that

$$\begin{aligned} J_p(K^*,B,A) &= \operatorname{Tr} \sqrt{A}\, K \, g_p\big(L_B R_A^{-1}\big)(K^*\sqrt{A}) \\ &= \operatorname{Tr} \sqrt{B} K^* \, \widetilde{g}_{1-p}\big(L_A R_B^{-1}\big)(\sqrt{B}) \\ &= \widetilde{J}_{1-p}(K,A,B) && (9) \end{aligned}$$

where, for $-1 \leq p < 1$, we define

$$\widetilde{g}_p(x) = x g_{1-p}(x^{-1}) = \begin{cases} \frac{1}{p(1-p)}(1-x^p) & p \neq 0 \\ -\log x & p = 0 \end{cases}. \tag{10}$$

and $\widetilde{J}_p(K,A,B) = \operatorname{Tr} \sqrt{B} K^* \, \widetilde{g}_p\big(L_A R_B^{-1}\big)(K\sqrt{B})$.

The functions $J_p(K,A,B)$ and $\widetilde{J}_p(K,A,B)$ have been considered before, usually with $K = I$, in the context of information geometry. (See [2, Section 7.2] and references therein.) What is novel here is that we present a simple unified proof of joint convexity in A, B that easily yields equality conditions, shows that they are independent of p and can be extended to other functions.

When $K = K^*$, the relation

$$J_p(K,A,A) = -\frac{1}{2p(1-p)} \operatorname{Tr}\,[K,A^p][K,A^{1-p}] \tag{11}$$

yields the original WYD information (up to a constant) and extends it to the range $(0,2]$. Morevoer, $K = K^*$ implies that $J_p(K,A,A) = J_{1-p}(K,A,A) = \widetilde{J}_p(K,A,A)$. Observe that although neither $g_p(w)$ nor $\widetilde{g}_p(w)$ is positive, their average $G_p(w) \equiv \frac{1}{2}[g(w) + wg(w^{-1}] \geq 0$ on $(0,\infty)$. Therefore, when $K = K^*$,

$$J_p(K,A,A) = \mathrm{Tr}\,(K\sqrt{A})^* G_p(L_A R_A^{-1})(K\sqrt{A}) \geq 0 \tag{12}$$

The function $J_p(I,A,B)$ is a more appealing generalization of relative entropy than $\mathrm{Tr}\, A^p B^{1-p}$ because of Proposition 1, which one can consider a generalization of Klein's inequality [16]. It allows one to use $J_p(I,A,B)$ as a pseudo-metric, as is commonly done with the relative entropy.

Proposition 1 *When U is unitary and $A, B > 0$ with $\mathrm{Tr}\, A = \mathrm{Tr}\, B = 1$, then $J_p(U,A,B) \geq 0$ with equality if and only if $A = U^* B U$.*

Proof: When U is unitary,

$$J_p(U,A,B) = J_p(I, UAU^*, B) = J_p(I, A, U^* B U). \tag{13}$$

Therefore, it suffices to consider the case $U = I$. For $p \in (0,1)$ Hölders inequality implies $\mathrm{Tr}\, A^p B^{1-p} \leq (\mathrm{Tr}\, A)^p (\mathrm{Tr}\, B)^{1-p} = 1$ with equality if and and only $A = B$. It immediately follows that

$$J_p(I,A,B) \geq \tfrac{1}{p(1-p)}\big(\mathrm{Tr}\, A - 1\big) = 0 \qquad \text{and} \qquad J_p(I,A,B) = 0 \Leftrightarrow A = B. \tag{14}$$

For $p = 1$, the result is well-known [37, Section 2.5.2] and originally due to O. Klein [16]. For $p \in (1,2)$ we write $p = 1 + r$ and again use Hölder's inequality

$$\begin{aligned} 1 &= \mathrm{Tr}\, A = \mathrm{Tr}\, A B^{-\frac{r}{r+1}} B^{\frac{r}{r+1}} \\ &\leq \|\big(A B^{-\frac{r}{r+1}}\|_{1+r}\, (\mathrm{Tr}\, B)^{\frac{r}{1+r}} \\ &\leq \big(\mathrm{Tr}\, A^{1+r} B^{-r}\big)^{\frac{1}{1+r}} \end{aligned} \tag{15}$$

where we used $\mathrm{Tr}\, B = 1$ and the second inequality follows from a classic result of Lieb-Thirring [24, Appendix B, Theorem 9] in the form given by Simon [38, Theorem 1.4.9]. **QED**

Because the denominator $p(1-p)$ changes sign at $p = 0$ and $p = 1$, both g and $\widetilde{g}$ are convex. In fact, they satisfy the much stronger condition of operator convexity for $p \in (0,2]$ and $p \in [-1,1)$ respectively.. Since $g(0) = 0$ and

$$\frac{g_p(x)}{x} = \begin{cases} \frac{1}{p(1-p)}(1 - x^{p-1}) & p \neq 1 \\ \log x & p = 1 \end{cases}, \tag{16}$$

it follows that $g_p(x)/x$ is operator monotone [3, 10, 26], for $p \in (0,2]$, i.e., g_p can be analytically continued to the upper half plane, which it maps into itself By applying Nevanlinna's theorem [1, Section 59, Theorem 2] to $g(x)/x$, one finds that $g(x)$ has an integral representation of the form

$$\begin{aligned} g_p(x) &= ax + \int_0^\infty \frac{x^2 t - x}{x+t}\, \nu(t)\, dt \\ &= ax + \int_0^\infty \Big[\frac{x^2}{x+t} - \frac{1}{t} + \frac{1}{x+t}\Big]\, t\nu(t)\, dt \end{aligned} \tag{17}$$

with $\nu(t) \geq 0$. Integral representations are not unique, and making a suitable change of variable in the classic formula

$$\int_0^\infty \frac{x^{p-1}}{x+1} = \frac{\pi}{\sin p\pi} \equiv \frac{1}{c_p} \qquad p \in (0,1) \tag{18}$$

allows us to give the following explicit representations

$$g_p(x) = \begin{cases} \frac{1}{p(1-p)}\Big[x + c_p \int_0^\infty \big(\frac{t}{x+t} - 1\big) t^{p-1} dt\Big] & p \in (0,1) \\ \int_0^\infty \big(\frac{x^2}{x+t} - 1 + \frac{t}{x+t}\big) \frac{1}{1+t} dt & p = 1 \\ \frac{1}{p(1-p)}\Big[x - c_{p-1} \int_0^\infty \frac{x^2}{x+t} t^{p-2} dt\Big] & p \in (1,2) \\ \frac{1}{2}(-x + x^2) & p = 2 \end{cases} \tag{19}$$

Note that for $p \in (0,2)$ the integrand is supported on $(0,\infty)$. This plays a key role in the equality conditions; therefore, we will henceforth concentrate on $p \in (0,2)$.

Theorem 2 *The function $J_p(K,A,B)$ defined in (9) is jointly convex in A, B.*

Proof: It follows from (17) that

$$\begin{aligned} J_p(K,A,B) &= a \mathrm{Tr}\, K^* A K \\ &+ \int_0^\infty \Big[\mathrm{Tr}\, K^* A \frac{1}{L_A + tR_B}(AK) - \frac{\mathrm{Tr}\, KBK^*}{t} + \mathrm{Tr}\, BK^* \frac{1}{L_A + tR_B}(KB)\Big] t\,\nu(t)\, dt \end{aligned} \tag{20}$$

The joint convexity then follows immediately from that of the map $(X, A, B) \mapsto \mathrm{Tr}\, X^* \frac{1}{L_A + tR_B}(X)$ which was proved in [36] following the strategy in [23]. The proof is also given in the Appendix. **QED**

For other approaches see Petz [29, 30], Effros [11], The advantage to the argument used here is that it immediately implies that equality holds in joint convexity if and only if it holds for each term in the integrand.

Corollary 3 *The relative entropy $H(A,B) = J_1(I,A,B)$ is jointly convex in A, B.*

2.2 Extensions with $r \neq 1 - p$.

We now consider extensions of Theorem 2 to situations in which B^{1-p} is replaced by B^r with $r \neq 1 - p$, using an idea from Bekjan [6] and Effros [11]. We will also show that equality holds in these extensions only under trivial conditions. For this we first need an elementary lemma.

Lemma 4 *Let $f(\lambda) : [0, \infty) \mapsto \mathbf{R}$ be a non-linear convex or concave operator function, let A_1, A_2 be density matrices and $A = \lambda A_1 + (1 - \lambda)A_2$ with $\lambda \in (0, 1)$. Then $f(A) = \lambda f(A_1) + (1 - \lambda)f(A_2)$ if and only if $A_1 = A_2$.*

Proof: Since any operator concave function is analytic, non-linearity implies that f is strictly concave. If $f(A) = \lambda f(A_1) + (1 - \lambda)f(A_2)$, then

$$\langle v, f(A)v \rangle = \lambda \langle v, f(A_1)v \rangle + (1 - \lambda)\langle v, f(A_2)v \rangle \tag{21}$$

for any vector v. Now choose v to be a normalized eigenvector of A. Then inserting this on the left above and applying Jensen's inequality to each term on the right, one finds

$$f\big(\langle v, Av \rangle\big) \leq \lambda f\big(\langle v, A_1 v \rangle\big) + (1 - \lambda)f\big(\langle v, A_2 v \rangle\big) \tag{22}$$

But this contradicts concavity unless equality holds, which implies that v is also an eigenvector of A_1 and A_2. But then the strict concavity of f also implies that $\langle v, A_1 v \rangle = \langle v, A_2 v \rangle$. Since this holds for an orthonormal basis of eigenvectors of A, we must have $A_1 = A_2$.

Corollary 5 *The function $(A, B) \mapsto \operatorname{Tr} K^* A^p K B^r$ is jointly concave on the set of positive definite matrices when $p, r \geq 0$ and $p + r \leq 1$. Moreover, when $p + r < 1$ and K is invertible, the convexity is strict unless $B_1 = B_2$ and $A_1 = A_2$.*

Proof: It is an immediate consequence of Theorem 2 that $(A, B) \mapsto \operatorname{Tr} K^* A^p K B^{1-p}$ is jointly concave in A, B. Now write $\operatorname{Tr} K^* A^p K B^r = \operatorname{Tr} K^* A^p K (B^s)^{1-p}$ with $s = r/(1 - p)$. First, observe that for $0 < s < 1$ the function $f(x) = x^s$ satisfies the hypotheses of Lemma 4. Therefore,

$$\big(\lambda B_1 + (1 - \lambda)B_2\big)^s > \lambda B_1^s + (1 - \lambda)B_2^s \tag{23}$$

with $0 < \lambda < 1$ and $B_1 \neq B_2$. The operator monotonicity of $x \mapsto x^{1-p}$ for $0 < p < 1$ then implies

$$\big(\lambda B_1 + (1 - \lambda)B_2\big)^r > \big(\lambda B_1^s + (1 - \lambda)B_2^s\big)^{1-p}, \tag{24}$$

and the joint concavity of $\mathrm{Tr}\, K^* A^p K B^{1-p}$ implies

$$\begin{aligned} \mathrm{Tr}\, K^* A^p K (B^s)^{1-p} &\geq \mathrm{Tr}\, K^* \big(\lambda A_1 + (1-\lambda) A_2\big)^p K \big(\lambda B_1^s + (1-\lambda) B_2^s\big)^{1-p} \\ &\geq \lambda \mathrm{Tr}\, K^* A_1^p K B_1^{s(1-p)} + (1-\lambda) \mathrm{Tr}\, K^* A_2^p K B_2^{s(1-p)} \end{aligned} \quad (25)$$

where $A = \lambda A_1 + (1-\lambda) A_2$, $B = \lambda B_1 + (1-\lambda) B_2$, which is precisely the joint concavity of $\mathrm{Tr}\, K^* A^p K B^r$. Moreover, equality in joint concavity implies equality in (25) and, since $K^* A^p K$ is strictly positive, this implies equality in (23). Therefore, equality in (25) gives a contradiction unless $B_1 = B_2$. In that case, the joint concavity reduces to concavity in A for which, by a similar argument, equality holds if and only if $A_1 = A_2$. **QED**

Corollary 6 *The function $(A, B) \mapsto \mathrm{Tr}\, K^* A^p K B^{1-r}$ is jointly convex on the set of positive definite matrices when $1 < r \leq p \leq 2$. Moreover, when $r < p$ and K is invertible, the convexity is strict unless $B_1 = B_2$ and $A_1 = A_2$.*

Proof: The argument is similar to that for Corollary 5. Write $\mathrm{Tr}\, K^* A^p K B^r = \mathrm{Tr}\, K^* A^p K (B^s)^{1-p}$ with $s = \frac{1-r}{1-p}$. Since $s \in (0, 1)$ and $1 - p \in (-1, 0)$ when $1 < r < p < 2$, it follows that x^s is operator concave and x^{1-p} is operator monotone decreasing. **QED**

2.3 Monotonicity under partial traces

Let X and Z denote the generalized Pauli operators whose action on the standard basis is $X|e_k\rangle = |e_{k+1}\rangle$ (with subscript addition mod d) and $Z|e_k\rangle = e^{i2\pi k/d}|e_k\rangle$. It is well known and easy to verify that $\frac{1}{d}\sum_k Z^k A Z^{-k}$ is the projection of a matrix onto the diagonal ones. If D is a diagonal matrix, then $\sum_k X^k D X^{-k} = (\mathrm{Tr}\, D) I$. Now let $\{W_n\}_{n=1,2\ldots d^2}$ denote some ordering of the generalized Pauli operators , e.g., $W_{j+k(d-1)} = X^j Z^k$ with $j, k = 1, 2 \ldots d$. Then $\frac{1}{d}\sum_n W_n A W_n^* = (\mathrm{Tr}\, A) I$ and

$$\tfrac{1}{d}\sum_n (W_n \otimes I_2)\, A_{12}\, (W_n \otimes I_2)^* = I_1 \otimes (\mathrm{Tr}_1\, A) = I_1 \otimes A_2 \quad (26)$$

Using the fact that replacing W_n by $U W_n U^*$ with U unitary, simply corresponds to a change of basis which does not affect (26) and then multiplying both sides by $U^* \otimes I_2$ on the left and $U \otimes I_2$ on the right gives the equivalent expression

$$\tfrac{1}{d}\sum_n (W_n U^* \otimes I_2)\, A_{12}\, (W_n U^* \otimes I_2)^* = I_1 \otimes A_2 \quad (27)$$

Combining this with joint convexity yields a slight generalization of the well-known monotonicity of relative entropy under partial traces (MPT), first proved by Lieb in [19] for the case $K_{12} = I_1 \otimes K_2$.

Theorem 7 *Let J_p be as in (7), A_{12}, B_{12} strictly positive in $M_{d_1} \otimes M_{d_2}$ and $K_{12} = V_1 \otimes K_2$ with V_1 unitary in M_{d_1}. Then*

$$J_p(K_2, A_2, B_2) \leq J_p(K_{12}, A_{12}, B_{12}) \tag{28}$$

Proof: Writing $\mathcal{W}_n$ for $W_n \otimes I_2$ and using (27) gives

$$\begin{aligned} J_p(K_2, A_2, B_2) &= \tfrac{1}{d_1} J_p(I_1 \otimes K_2, I_1 \otimes A_2, I_1 \otimes B_2) \\ &= \tfrac{1}{d_1} J_p\Big(I_1 \otimes K_2,\ \tfrac{1}{d_1} \textstyle\sum_n \mathcal{W}_n (V_1^* \otimes I_2) A_{12} (V_1 \otimes I_2) \mathcal{W}_n^*\ \tfrac{1}{d_1} \sum_n \mathcal{W}_n B_{12} \mathcal{W}_n^*\Big) \\ &\leq \tfrac{1}{d_1^2} \sum_n J_p\big(I_1 \otimes K_2, \mathcal{W}_n (V_1^* \otimes I_2) A_{12} (V_1 \otimes I_2) \mathcal{W}_n^*, \mathcal{W}_n B_{12} \mathcal{W}_n^*\big) \\ &= J_p(V_1 \otimes K_2, A_{12}, B_{12}) \end{aligned}$$

where the final equality follows from the unitary invariance of the trace. **QED**

Because $\mathrm{Tr}_{12}(V_1 \otimes K_2) A_{12} (V_1 \otimes K_2)^* = \mathrm{Tr}_2 K_2 A_2 K_2^*$, (28) is equivalent to

$$\mathrm{Tr}\, K_2^* A_2^p K_2 B_2^{1-p} - \mathrm{Tr}\,(V_1 \otimes K_2)^* A_{12}^p (V_1 \otimes K_2) B_{12}^{1-p} \begin{cases} \geq 0 & p \in (0,1) \\ \leq 0 & p \in (1,2) \end{cases}. \tag{29}$$

We can obtain a weak reversal of this for $p \in (0,1)$. The argument in the Appendix shows that for any p and fixed $A, B \geq 0$ both $\mathrm{Tr}\, K^* A^p K B^{1-p}$ and $\mathrm{Tr}\, K^* A K$ are convex in K. This was observed earlier by Lieb [19] and also follows from the results in [23]. One can then apply the argument above in the special case $A_{12} = I_1 \otimes A_2, B_{12} = I_1 \otimes B_2$ to conclude that

$$\begin{aligned} \mathrm{Tr}\, K_2^* A_2^p K_2 B_2^{1-p} &\leq \tfrac{1}{d_1} \mathrm{Tr}\, K_{12}^* (I_1 \otimes A_2)^p K_{12} (I_1 \otimes B_2)^{1-p} & (30) \\ &\leq \mathrm{Tr}\, K_{12}^* (I_1 \otimes A_2)^p K_{12} (I_1 \otimes B_2)^{1-p} & (31) \end{aligned}$$

independent of whether $p < 1$ or $p > 1$. However, because the term $\mathrm{Tr}\, K^* A K$ is convex rather than linear in K, (30) does not allow us to draw any conclusions about the monotonicity of $J_p(K_{12}, I_1 \otimes A_2, I_1 \otimes B_2)$.

To prove Theorem 2.3 we showed that joint convexity implies monotonicity; the reverse implication also holds. Let $A_1, \ldots, A_m$, $B_1, \ldots, B_m$ be positive definite matrices in M_d, $A = \sum_j A_j$, $B = \sum_j B_j$, and put

$$\widetilde{A}_{12} = \sum_j |e_j\rangle\langle e_j| \otimes A_j, \quad \widetilde{B}_{12} = \sum_j |e_j\rangle\langle e_j| \otimes B_j, \tag{32}$$

for $e_1, \ldots, e_m$ the standard basis of $\mathbf{C}^m$. Then $\widetilde{A}_{12}$ and $\widetilde{B}_{12}$ are block diagonal, and $\widetilde{A}_2 = \mathrm{Tr}_1 \widetilde{A}_{12} = \sum_k A_k = A$ and similarly for B. Then if monotonicity under partial

traces holds, one can conclude that

$$\begin{aligned} J_p(K,A,B) &= J_p(K_2,\widetilde{A}_2,\widetilde{B}_2) \\ &\leq J_p(I_1\otimes K,\widetilde{A}_{12},\widetilde{B}_{12}) = \sum_j J_p(K,A_j,B_j) \end{aligned} \tag{33}$$

Thus, monotonicity under partial traces also directly implies joint convexity of J_p.

Applying (28) in the case $K=I$, and $A_{12}\mapsto A_{123}$ and $B_{12}\mapsto A_{12}\otimes I_3$ gives

$$J_p(I_{12},A_{23},A_2\otimes I_3) \leq J_p(I_{123},A_{123},A_{12}\otimes I_3) \tag{34}$$

When $p=1$, it follows from (7) that

$$J_1(I_{12},A_{23},A_2\otimes I_3) = H(A_{23},A_2\otimes I_2) = -S(A_{23})+S(A_2)$$

where $S(A)=-\mathrm{Tr}\,A\log A$. Thus, (34) becomes

$$-S(A_{23})+S(A_2)\leq -S(A_{123})+S(A_{12})$$

or, equivalently

$$S(A_2)+S(A_{123})\leq S(A_{12})+S(A_{23}) \tag{35}$$

which is the standard form of SSA.

3 Equality for joint convexity of $J_p(K,A,B)$.

3.1 Origin of necessary and sufficient conditions

Looking back at the proof of Theorem 2, we see that for $p\in(0,2)$, equality holds in the joint convexity of $J_p(K,A,B)$ if and only if equality holds in the joint convexity for each term in the integrand in (17). It should be clear from the argument given in the Appendix, that this requires $M_j=0$ for all j with M_j given by (70). This is easily seen to be equivalent to

$$(L_{A_j}+tR_{B_j})^{-1}(X_j) = (L_A+tR_B)^{-1}(X) \quad \text{for all } j. \tag{36}$$

with $X_j=A_jK$ and/or $X_j=KB_j$. By writing $AK=L_A(K)$ in the former case and $KB=R_B(K)$ in the latter we obtain the conditions

$$(I+t\Delta_{A_jB_j}^{-1})^{-1}(K) = (I+t\Delta_{AB}^{-1})^{-1}(K) \quad \forall\, j \quad \forall t>0 \tag{37a}$$

$$(\Delta_{A_jB_j}+tI)^{-1}(K) = (\Delta_{AB}+tI)^{-1}(K) \quad \forall\, j \quad \forall t>0 \tag{37b}$$

From the integral representations (19), one might expect it to be necessary for either or both of (37a) and (37b) to hold depending on p. In fact, either will suffice because (37a) holds if and only if (37b) holds. Because Δ_{AB} is positive definite, by analytic continuation (37b) extends from $t > 0$ to the entire complex plane, except points $-t$ on the negative real axis for which $t \in \text{spectrum}(\Delta_{AB})$. Therefore, by using the Cauchy integral formula, one finds that for any function G analytic on $\mathbf{C}\backslash(-\infty, 0]$ $G(\Delta_{A_jB_j})(K) = G(\Delta_{AB})(K)$.

Theorem 8 *For fixed K, and $A = \sum_j A_j, B = \sum_j B_j$, the following are equivalent*

a) $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ *for all* $p \in (0, 2)$.

b) $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ *for some* $p \in (0, 2)$.

c) $(\Delta_{A_jB_j} + tI)^{-1}(K) = (\Delta_{AB} + tI)^{-1}(K)$ *for all* j *and for all* $t > 0$.

d) $A_j^{it} K B_j^{-it} = A^{it} K B^{-it}$ *for all* j *and for all* $t > 0$.

e) $(\log A - \log A_j)K = K(\log B - \log B_j)$ *for all* j.

Proof: Clearly (a) $\Rightarrow$ (b). The implications (b) $\Rightarrow$ (c) $\Rightarrow$ (d), as well as (b) $\Rightarrow$ (a), follow from the discussion above. Differentiation of (d) at $t = 0$ gives (d) $\Rightarrow$ (e), and it is straightforward to verify that (e) $\Rightarrow$ (b) with $p = 1$. Moreover, (d) implies $\sum_j \operatorname{Tr} K^* A_j^{it} K B_j^{1-it} = \operatorname{Tr} K^* A^{it} K B^{1-it}$ for all t, which implies (a) by analytic continuation. **QED**

3.2 Sufficient subalgebras

When $K = I$, we can obtain a more useful reformulation of the equality conditions by using results about sufficient subalgebras obtained in [14, 15, 32]. Since the definition and convexity properties of $J_p(I, A, B)$ extend by continuity to positive semidefinite matrices, with $\ker B \subseteq \ker A$, we will formulate the conditions in this more general situation, using the conventions in Section 1.2.

Let $N \subseteq M_d$ be a subalgebra, then there is a trace preserving conditional expectation E_N from M_d onto N, such that $\operatorname{Tr} AX = \operatorname{Tr} E_N(A)X$ for all $X \in N$. In particular, if $N = M_{d_1} \otimes I \subseteq M_{d_1} \otimes M_{d_2}$, then we have $E_N(A_{12}) = \operatorname{Tr}_2 A \otimes \frac{1}{d_2} I$.

Let $Q_1, \ldots, Q_m \in M_d^+$ and assume that $\ker Q_m \subseteq \ker Q_j$ for all j. The subalgebra N is said to be sufficient for $\{Q_1, \ldots, Q_m\}$ if there is a completely positive trace preserving map $T : N \to M_d$, such that $T(E_N(Q_j)) = Q_j$ for all $j = 1, \ldots, m$. This definition is due to Petz [32, 31] and it is a quantum generalization of the well known notion of sufficiency from classical statistics. In [32], it was shown that sufficient subalgebras can be characterized by the condition

$$H(Q_j, Q_m) = H(E_N(Q_j), E_N(Q_m)), \quad \text{for all } j$$

We combine this with the results of the previous section to obtain other useful characterizations of sufficiency.

Theorem 9 *Let $Q_1, \ldots, Q_m \in M_d^+$ be such that $\ker Q_m \subseteq \ker Q_j$ for all j. Let $N \subseteq M_d$. The following are equivalent.*

(i) N is sufficient for $\{Q_1, \ldots, Q_m\}$.

(ii) $E_N(Q_j)^{it} E_N(Q_m)^{-it} P_{(\ker Q_m)^\perp} = Q_j^{it} Q_m^{-it}$, for all j, $t \in \mathbf{R}$.

(iii) There exist $Q_{j,0} \in N^+$, and $D \in M_d^+$, such that $\ker D = \ker Q_m$, and $Q_j = Q_{j,0} D$ for $j = 1, \ldots, m$.

(iv) $J_p(I, Q_j, Q_m) = J_p(I, E_N(Q_j), E_N(Q_m))$ for all j and some $p \in (0,1)$

The proof of the conditions (i) – (iii) can be found in [14], see also [27]. The condition (iv) was proved in [15].

3.3 Equality conditions with $K = I$

Theorem 10 *Let $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ be positive semi-definite matrices with $\ker B_j \subseteq \ker A_j$, and let $A = \sum_j A_j, B = \sum_j B_j$. Then the following are equivalent.*

a) $J_p(I, A, B) = \sum_j J_p(I, A_j, B_j)$ for all $p \in (0,2)$.

b) $J_p(I, A, B) = \sum_j J_p(I, A_j, B_j)$ for some $p \in (0,2)$.

c) $A_j^{it} B_j^{-it} = A^{it} B^{-it} P_{(\ker B_j)^\perp}$ for all j and $t \in \mathbf{R}$

d) There are positive matrices $D_1, \ldots, D_m$, with $\ker D_j = \ker B_j$, such that $[A_j, D_j] = [B_j, D_j] = 0$, and with $D = \sum_j D_j$

$$A_j = AD^{-1}D_j, \qquad B_j = BD^{-1}D_j \tag{38}$$

Proof: As in Section 3.1, (b) implies (36) on $(\ker B_j)^\perp$, with $X_j = B_j, X = B$. This gives

$$(\Delta_{A_j B_j} + tI)^{-1}(I) = (\Delta_{AB} + tI)^{-1}(I) \qquad \text{on} \quad (\ker B_j)^\perp. \tag{39}$$

Then (c) follows from the Cauchy integral formula as in Section 3.1.

To show (c) implies (d), we will use Theorem 9. First let $N = I \otimes M_d \subseteq M_m \otimes M_d$ and let $\widetilde{A}_{12}$, $\widetilde{B}_{12}$ be the block-diagonal matrices in $M_m \otimes M_d$, defined by (32). Clearly, we have $\ker \widetilde{A}_{12} \supseteq \ker \widetilde{B}_{12} = \sum_j |e_j\rangle\langle e_j| \otimes \ker B_j$ and $E_N(\widetilde{A}_{12}) = \frac{1}{m} I \otimes A$, $E_N(\widetilde{B}_{12}) = \frac{1}{m} I \otimes B$. Then (c) implies $E_N(\widetilde{A}_{12})^{it} E_N(\widetilde{B}_{12})^{-it} P_{(\ker \widetilde{B}_{12})^\perp} = \widetilde{A}_{12}^{it} \widetilde{B}_{12}^{-it}$ for all t. Then by using Theorem 9 with $Q_1 = \widetilde{A}_{12}, Q_m = Q_2 = \widetilde{B}_{12}$, we can conclude that there are

positive matrices $A_0, B_0 \in M_d$ and $D_{12} \in (M_m \otimes M_d)^+$, such that $\ker D_{12} = \ker \widetilde{B}_{12}$, $[I \otimes A_0, D_{12}] = [I \otimes B_0, D_{12}] = 0$ and

$$\widetilde{A}_{12} = (I \otimes A_0) D_{12}, \qquad \widetilde{B}_{12} = (I \otimes B_0) D_{12} \tag{40}$$

Since $\widetilde{A}_{12}$, $\widetilde{B}_{12}$ are block diagonal, $D_{12} = \sum_j |e_j\rangle\langle e_j| \otimes D_j$ must also be block diagonal with $D_j \in M_d^+$, $\ker D_j = \ker B_j$, $[A_0, D_j] = [B_0, D_j] = 0$ for all j and

$$A_j = A_0 D_j, \qquad B_j = B_0 D_j. \tag{41}$$

Taking Tr_1 in (40) gives $A = A_0 D$ and $B = B_0 D$. Using this in (41) gives (38) which proves (d). The implications (d) $\Rightarrow$ (a) $\Rightarrow$ (b) are straightforward. **QED**

We return briefly to the case of arbitrary K. Note that if the condition (d) holds and $[D_j, K] = 0$ for all j, then $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ for all $p \in (0, 2)$, this gives a sufficient, but not necessary, condition for equality if $K \neq I$. The next result reduces the case of K unitary to $K = I$. Then, we can apply the conditions of Theorem 10 to A_j and $K B_j K^*$.

Theorem 11 *If K is unitary, then $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ if and only if $J_p(I, A, KBK^*) = \sum_j J_p(I, A_j, K B_j K^*)$*

Proof: When K is unitary, then $K B^p K^* = (KBK^*)^p$ which implies $J_p(K, A, B) = J_p(I, A, KBK^*)$. **QED**

One can try to extend the results of this section to the case $\|K\| \leq 1$, and hence to all K, by using the unitary dilation

$$\mathcal{U} = \begin{pmatrix} K & L \\ -L & K \end{pmatrix}$$

where $L = U(1 - |K|^2)^{1/2}$ and $K = U|K|$ is the polar decomposition. Then, with

$$\mathcal{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

we have $J_p(K, A, B) = J_p(\mathcal{U}, \mathcal{A}, \mathcal{B})$, so that we may use Theorem 11 to get conditions for equality. But note that the conditions of Theorem 10 require that $\ker \mathcal{U}\mathcal{B}_j\mathcal{U}^* \subseteq \ker \mathcal{A}_j$ and it can be shown that this implies $P_{(\ker A_j)^\perp} K P_{(\ker B_j)^\perp} K^* = P_{(\ker A_j)^\perp}$, where $P_{\mathcal{N}}$ denotes a projection onto the subscripted space. In particular, if all A_j and B_j are invertible, this restricts us to unitary K.

3.4 Equality in monotonicity under partial trace

It is easy to see that when $A_{12} = A_1 \otimes A_2$ and $B_{12} = B_1 \otimes B_2$, then $J_p(I, A_{12}, B_{12}) = J_p(I, A_2, B_2)$ if and only if $A_1 = B_1$ with $\operatorname{Tr} A_1 = 1$. However, it is not necessary that $A_{12} = A_1 \otimes A_2$. The equality conditions are given by the following theorem.

Theorem 12 *Let $K = I$ and $A_{12}, B_{12} \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)^+$, with $\ker B_{12} \subseteq \ker A_{12}$. Equality holds in (28) if and only if*

(i) $\mathcal{H}_2 = \bigoplus_n \mathcal{H}_n^L \otimes \mathcal{H}_n^R$.

(ii) $A_{12} = \bigoplus_n A_n^L \otimes A_n^R$ with $A_n^L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)^+$ and $A_n^R \in \mathcal{B}(\mathcal{H}_n^R)^+$

(iii) $B_{12} = \bigoplus_n B_n^L \otimes B_n^R$ with $B_n^L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)^+$ and $B_n^R \in \mathcal{B}(\mathcal{H}_n^R)^+$

(iv) $A_n^L = B_n^L$ for all n

Proof: Let us denote $A_j = \frac{1}{d_1}\mathcal{W}_j A_{12} \mathcal{W}_j^*$, $B_j = \frac{1}{d_1}\mathcal{W}_j B_{12} \mathcal{W}_j^*$, with $\mathcal{W}_j$ defined as in the proof of Theorem 7. Then we get that equality in (28) is equivalent to

$$J_p(I_{12}, \sum_j A_j, \sum_j B_j) = \sum_j J_p(I_{12}, A_j, B_j)$$

By Theorem 10, equality for some p implies equality for all p, so that $J_p(I, A_{12}, B_{12}) = J_p(I, \operatorname{Tr}_1 A, \operatorname{Tr}_1 B) = J_p(I, E_N(A_{12}), E_N(B_{12}))$ for $p \in (0, 1)$, where N is the subalgebra $I_1 \otimes \mathcal{B}(\mathcal{H}_2) \subseteq \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. Hence N is sufficient for $\{A_{12}, B_{12}\}$ and, by Theorem 9, there are some $A_R, B_R \in \mathcal{B}(\mathcal{H}_2)^+$ and $D \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)^+$, $\ker D = \ker B_{12}$, such that $[(I_1 \otimes A_R), D] = [(I_1 \otimes B_R), D] = 0$ and

$$A_{12} = D(I_1 \otimes A_R), \quad B_{12} = D(I_1 \otimes B_R) \tag{42}$$

Now let M_1 be the subalgebra in $\mathcal{B}(\mathcal{H}_2)$, generated by A_R, B_R. Then $D \in (I_1 \otimes M_1)' = \mathcal{B}(\mathcal{H}_1) \otimes M_1'$ where M' denotes the commutant of M. There is a decomposition $\mathcal{H}_2 = \bigoplus_n \mathcal{H}_n^L \otimes \mathcal{H}_n^R$, such that

$$M_1' = \bigoplus_n \mathcal{B}(\mathcal{H}_n^L) \otimes 1_n^R, \qquad M_1 = \bigoplus_n 1_n^L \otimes \mathcal{B}(\mathcal{H}_n^R)$$

and $D = \bigoplus_n D_n \otimes 1_n^R$, where $D_n \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)$. Since $A_R, B_R \in M_1$, we get the result, with $A_n^L = B_n^L = D_n$. The converse is can be verified directly **QED**

Applying this result in the case $A_{12} \mapsto A_{123}$ and $B_{12} \mapsto A_{12} \otimes I_3$ gives equality conditions in (34). Since these are independent of p, they are identical to the conditions, first given in [13], for equality in SSA (35) which corresponds to $p = 1$.

Corollary 13 *Equality holds in (34) if and only if*

(i) $\mathcal{H}_2 = \bigoplus_n \mathcal{H}_n^L \otimes \mathcal{H}_n^R$.

(ii) $A_{123} = \bigoplus_n A_n^L \otimes A_n^R$ with $A_n^L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)$ and $A_n^R \in \mathcal{B}(\mathcal{H}_n^R \otimes \mathcal{H}_3)$

Proof: It suffices to let $A_{12} \to A_{123}$ and $B_{12} \to A_{12} \otimes I_3$ in Theorem 12. **QED**

To apply these results in Section 4, it is useful to observe that condition (ii) in Corollary 13 above can be written as

$$A = (F_L \otimes I_3)(I_1 \otimes F_R) \tag{43}$$

with $F_L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)^+$, $F_R \in \mathcal{B}(\mathcal{H}_2 \otimes \mathcal{H}_3)^+$, $[F_L \otimes I_3, I_1 \otimes F_R] = 0$. Combining this with part (d) of Theorem 10 gives the following useful result, which essentially allows us to bypass the need to apply Theorem 10 to $J_p(I, A_j, \mathcal{W}_n A_j \mathcal{W}_n)$.

Corollary 14 *Let $A_j \in M_{d_1} \otimes M_{d_2}$, $A = \sum A_j$. Then*

$$J_p(I_{12}, A, (\mathrm{Tr}_2 A) \otimes I_2) = \sum_j J_p(I_{12}, A_j, (\mathrm{Tr}_2 A_j) \otimes I_2) \tag{44}$$

if and only if there are $D_j \in M_{d_1}^+$, such that $\ker D_j = \ker \mathrm{Tr}_2 A_j$, $[A_j, D_j \otimes I] = 0$ *and* $A_j = A(D^{-1} D_j \otimes I)$ *with* $D = \sum_j D_j$.

Proof: Let $\widetilde{A}_{123} = \sum_j |e_j\rangle\langle e_j| \otimes A_j \in M_m \otimes M_{d_1} \otimes M_{d_2}$, then $A = \widetilde{A}_{23} \in M_{d_1} \otimes M_{d_2}$ and (44) can be written as

$$J_p(I_{23}, \widetilde{A}_{23}, \widetilde{A}_2 \otimes I_3) = J_p(I_{123}, \widetilde{A}_{123}, \widetilde{A}_{12} \otimes I_3)$$

By (43), this is equivalent to the existence of F_L and F_R, $[(F_L \otimes I_3), (I_1 \otimes F_R)] = 0$, such that $\widetilde{A}_{123} = (F_L \otimes I_3)(I_1 \otimes F_R)$. Since $\widetilde{A}_{(1)(23)}$ is block-diagonal, F_L must be of the form $F_L = \sum_j |e_j\rangle\langle e_j| \otimes D_j$, so that $A_j = F_R(D_j \otimes I)$. But taking Tr_1 of (43) gives $A = (D \otimes I_3) F_R = F_R(D \otimes I_3)$ so that $A_j = A(D^{-1} D_j \otimes I)$. **QED**

4 Equality in joint convexity of Carlen-Lieb

Carlen and Lieb [8] obtain several convexity inequalities from those of the map

$$\Upsilon_{p,q}(K, A) \equiv \mathrm{Tr}\,(K^* A^p K)^{q/p}. \tag{45}$$

using an identity which we write only for $q = 1$ and $p > 1$ in our notation as

$$\Upsilon_{p,1}(K, A) = (p-1) \inf \big\{ J_p(K, A, X) + \tfrac{1}{p} \mathrm{Tr}\, X + \tfrac{1}{p(p-1)} \mathrm{Tr}\, K^* A K : X > 0 \big\} \tag{46}$$

We introduce the closely related quantity

$$\begin{aligned} \widehat{\Upsilon}_{p,1}(K, A) &= \inf \big\{ J_p(K, A, X) + \tfrac{1}{p} \mathrm{Tr}\, X : X > 0 \big\} && (47) \\ &= \tfrac{1}{(p-1)} \big(\Upsilon_{p,1}(K, A) - \tfrac{1}{p} \mathrm{Tr}\, K^* A K \big) && (48) \end{aligned}$$

which is well-defined for all $p \in (0,2)$ and allows us to continue to treat the cases $p < 1$ and $p > 1$ simultaneously, as well as include the special case $p = 1$ for which

$$\begin{aligned} \widehat{\Upsilon}_{1,1}(K,A) &= -\mathrm{Tr}\, K^*AK \log(K^*AK) + \mathrm{Tr}\, K^*(A\log A)K \\ &= S(K^*AK) + \mathrm{Tr}\, KK^*A\log A \end{aligned} \tag{49}$$

Since we are dealing with finite dimensional spaces, the infimum in (46) has a minimizer which satisfies

$$X_{\min} = (K^*A^pK)^{1/p}. \tag{50}$$

For fixed K, let X_j denote the minimizer associated with A_j. Then

$$\begin{aligned} \widehat{\Upsilon}_{p,1}(K,A_1) + \widehat{\Upsilon}_{p,1}(K,A_2) &= J_p(K,A_1,X_1) + \tfrac{1}{p}\mathrm{Tr}\, X_1 + J_p(K,A_2,X_2) + \tfrac{1}{p}\mathrm{Tr}\, X_2 \\ &\geq J_p(K,A_1+A_2,X_1+X_2) + \tfrac{1}{p}\mathrm{Tr}\,(X_1+X_2) \qquad (51) \\ &\geq \inf\left\{J_p(K,A_1+A_2,X) + \tfrac{1}{p}\mathrm{Tr}\, X : X > 0\right\} \\ &= \widehat{\Upsilon}_{p,1}(K,A_1+A_2) \qquad (52) \end{aligned}$$

which proves convexity of $\widehat{\Upsilon}_{p,1}$. Note that equality above requires both $X = \sum_j X_j$ and $J_p(K,A,X) = \sum_j J_p(K,A_j,X_j)$. where X is the minimizer associated with A.

Now we introduce some notation following the strategy in the published version of [8]. Let $|\mathbb{1}\rangle$ denote the vector $(1,1,\ldots,1)$ with all components 1 and $|e_1\rangle$ the vector $(1,0,\ldots,0)$. Define

$$\mathcal{K} = \tfrac{1}{d} I \otimes |\mathbb{1}\rangle\langle e_1| = \begin{pmatrix} I & 0 & \ldots & 0 \\ I & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ I & 0 & \ldots & 0 \end{pmatrix} \tag{53}$$

and

$$\mathcal{A}_j = \sum_k A_{jk} \otimes |e_k\rangle\langle e_k| = \begin{pmatrix} A_{j1} & 0 & 0 & \ldots & 0 \\ 0 & A_{j2} & 0 & \ldots & 0 \\ 0 & 0 & A_{j3} & \ldots & 0 \\ \vdots & & & \ddots & \vdots \end{pmatrix}, \tag{54}$$

and $\mathcal{A} = \sum_j \mathcal{A}_j = \sum_k A_k \otimes |e_k\rangle\langle e_k|$ with $A_k = \sum_j A_{jk}$. Then

$$\mathcal{K}^*\mathcal{A}^p\mathcal{K} = \big(\sum_k A_k^p\big) \otimes |e_1\rangle\langle e_1|$$

With this notation, we make some definitions following Carlen and Lieb but modified to allow a unified treatment of $p \in (0,2)$.

$$\begin{aligned}
\Phi_{(p,1)}(\mathcal{A}) = \Phi_{(p,1)}(A_1, A_2, A_3 \ldots) &\equiv \operatorname{Tr}\left(A_1^p + A_2^p + A_3^p + \ldots\right)^{1/p} && (55)\\
&= \Upsilon_{p,1}(\mathcal{K}, \mathcal{A})\\
\widehat{\Phi}_{(p,1)}(\mathcal{A}) = \widehat{\Phi}_{(p,1)}(A_1, A_2, A_3 \ldots) &\equiv \widehat{\Upsilon}_{p,1}(\mathcal{K}, \mathcal{A}) && (56)\\
&= \tfrac{1}{(p-1)}\Big[\Phi_{(p,1)}(A_1, A_2, A_3 \ldots) - \tfrac{1}{p}\sum_k \operatorname{Tr} A_k\Big]
\end{aligned}$$

The definitions of Φ and $\widehat{\Phi}$ apply only when $\mathcal{A}$ is a block diagonal matrix in $M_{d_1} \otimes M_{d_2}$. We now extend this to an arbitrary matrices $\mathcal{A}_{12} \in M_{d_1} \otimes M_{d_2}$.

$$\begin{aligned}
\Psi_{(p,1)}(\mathcal{A}_{12}) &\equiv \operatorname{Tr}_1\left(\operatorname{Tr}_2 \mathcal{A}_{12}^p\right)^{1/p} && (57)\\
\widehat{\Psi}_{(p,1)}(\mathcal{A}_{12}) &\equiv \tfrac{1}{(p-1)}\left[\Psi_{(p,1)}(\mathcal{A}_{12}) - \tfrac{1}{p}\operatorname{Tr} \mathcal{A}_{12}\right] && (58)
\end{aligned}$$

For $p = 1$, the formulas with hats reduce to the conditional entropy

$$\begin{aligned}
\widehat{\Phi}_{(1,1)}(A_1, A_2, A_3 \ldots) &= -\operatorname{Tr}\left(\textstyle\sum_j A_j\right)\log\left(\textstyle\sum_j A_j\right) + \textstyle\sum_j A_j \log A_j\\
&= S\big(\textstyle\sum_j A_j\big) - S(\mathcal{A}_{12}) = J_1\big(I, \mathcal{A}_{12}, \operatorname{Tr}_2 \mathcal{A}_{12} \otimes I_2\big) && (59)\\
\widehat{\Psi}_{(1,1)}(\mathcal{A}_{12}) &= S(\mathcal{A}_1) - S(\mathcal{A}_{12}) = H(\mathcal{A}_{12}, \mathcal{A}_1 \otimes I_2) && (60)
\end{aligned}$$

When $\mathcal{A}_{12}$ is block diagonal, $\Psi_{(p,1)}(\mathcal{A}_) = \Phi_{(p,1)}(\mathcal{A})$ with the understanding that $\operatorname{Tr}_2\mathcal{A} = \sum_k A_k$. Now let W_n denote the generalized Pauli matrices as in Section 2.3, $\mathcal{W}_n = I_1 \otimes W_n$ and define

$$\mathcal{A}_{123} = \sum_n \mathcal{W}_n \mathcal{A} \mathcal{W}_n^* \otimes |e_n\rangle\langle e_n| \tag{61}$$

so that $\mathcal{A}_{123}$ is block diagonal with blocks $\mathcal{W}_n \mathcal{A}_{12} \mathcal{W}_n^*$. Then

$$d_2^{\frac{1+p}{p}}\Psi_{(p,1)}(\mathcal{A}_{12}) = \Phi(\mathcal{A}_{(12)(3)}) = \Phi\big(\mathcal{W}_1\mathcal{A}_{12}\mathcal{W}_1^*,\, \mathcal{W}_2\mathcal{A}_{12}\mathcal{W}_2^*,\, \ldots\big). \tag{62}$$

It is straightforward to show that for $p \in (0,2)$ the functions $\widehat{\Phi}_{(p,1)}(\mathcal{A})$ and $\widehat{\Psi}_{(p,1)}(\mathcal{A})$ are all convex in $\mathcal{A}$, inheriting this property from the quantities from which they are defined. In view of (59) and (60), the conditions for equality in the next two theorems are not surprising.

Theorem 15 *The function $\widehat{\Phi}_{(p,1)}(\mathcal{A})$ is convex in $\mathcal{A}$ for $p \in (0,2)$. Moreover, the following are equivalent*

(i) $J_p\big(I, \mathcal{A}, (\operatorname{Tr}_2 \mathcal{A}) \otimes I_2\big) = \sum_j J_p\big(I, \mathcal{A}_j, (\operatorname{Tr}_2 \mathcal{A}_j) \otimes I_2\big)$

(ii) There are matrices $D_j > 0$, $D = \sum_j D_j$, such that $[A_{jk}, D_j] = 0$, $\ker D_j = \ker(\sum_k A_{jk})$ and $A_{jk} = A_k D^{-1} D_j$.

(iii) $\widehat{\Phi}_{(p,1)}(A_1, A_2, A_3 \ldots) = \sum_j \widehat{\Phi}_{(p,1)}(A_{j1}, A_{j2}, A_{j3} \ldots)$

Proof: It follows from Corollary 14 and the fact that $\mathcal{A}_j$ are block-diagonal that (i) $\Leftrightarrow$ (ii) and it is straightforward to verify that (ii) $\Rightarrow$ (iii). Moreover, (iii) implies (i) for $p=1$, by (59). To show that (iii) implies (ii) for $p \neq 1$, observe that (iii), implies $\widehat{\Upsilon}_{p,1}(\mathcal{K},\mathcal{A}) = \sum_j \widehat{\Upsilon}_{p,1}(\mathcal{K},\mathcal{A}_j)$, and this implies

$$J_p(\mathcal{K},\mathcal{A},\mathcal{X}) = \sum_j J_p(\mathcal{K},\mathcal{A}_j,\mathcal{X}_j) \tag{63}$$

where $\mathcal{X}_j = (\mathcal{K}^*\mathcal{A}_j^p\mathcal{K})^{1/p} = X_j \otimes |e_1\rangle\langle e_1|$ and $\sum_j \mathcal{X}_j = \mathcal{X} = (\mathcal{K}^*\mathcal{A}^p\mathcal{K})^{1/p} = X \otimes |e_1\rangle\langle e_1|$, with $X_j = (\sum_k A_{jk}^p)^{1/p}$ and $X = (\sum_k A_k^p)^{1/p}$. Since

$$\mathcal{K}^*\mathcal{A}_j^p\mathcal{K}\mathcal{X}_j^{1-p} = \sum_k A_{jk}^p X_j^{1-p} \otimes |e_1\rangle\langle e_1|,$$

with a similar expression for $\mathcal{K}^*\mathcal{A}^p\mathcal{K}\mathcal{X}^{1-p}$, we find

$$\sum_k J_p(I,A_k,X) = J_p(\mathcal{K},\mathcal{A},\mathcal{X}) = \sum_j J_p(\mathcal{K},\mathcal{A}_j,\mathcal{X}_j) = \sum_{k,j} J_p(I,A_{jk},X_j)$$

Convexity then implies that we must have

$$J_p(I,A_k,X) = \sum_j J_p(I,A_{jk},X_j) \qquad \forall\ k. \tag{64}$$

Since $\ker X_j \subseteq \ker A_{jk}$, Theorem 10 implies that

$$A_k^{it}X^{-it}P_{(\ker X_j)^\perp} = A_{jk}^{it}X_j^{-it}, \quad \text{ for all } k,j,t \tag{65}$$

After writing $\widetilde{A}_k = \sum_j |e_j\rangle\langle e_j| \otimes A_{jk}$, $\widetilde{X} = \sum_j |e_j\rangle\langle e_j| \otimes X_j$, this reads

$$\widetilde{A}_k^{it}\widetilde{X}^{-it} = (\tfrac{1}{m}I \otimes \mathrm{Tr}_1\widetilde{A}_k)^{it}(\tfrac{1}{m}I \otimes \mathrm{Tr}_1\widetilde{X})^{-it}P_{(\ker \widetilde{X})^\perp},$$

so that, by Theorem 10, there are elements $B_k \in M_d^+$ and $D \in (M_m \otimes M_d)^+$, such that $[(I \otimes B_k), D] = 0$ and $\widetilde{A}_k = (I \otimes B_k)D$. As before, one finds $D = \sum_j |e_j\rangle\langle e_j| \otimes D_j$ for some $D_j \in M_d^+$ which implies (ii). **QED**

Theorem 16 *The function $\widehat{\Psi}_{(p,1)}(\mathcal{A}_{12})$ is convex in $\mathcal{A}_{12}$ for $p \in (0,2)$. Moreover, if we let $\mathcal{A}_{123}$ denote the block diagonal matrix with blocks $\mathcal{W}_n\mathcal{A}\mathcal{W}_n^*$, the following are equivalent*

(i) $J_p(I,\mathcal{A}_{123},\mathcal{A}_1 \otimes I_{23}) = \sum_j J_p\big(I,(\mathcal{A}_{123})_j,(\mathcal{A}_1)_j \otimes I_{23}\big)$ with $\mathcal{A}_{123}$ defined by (62).

(ii) There are matrices $D_j \in M_{d_1}^+$, $D = \sum_j D_j$, such that $[\mathcal{A}_j, D_j \otimes I] = 0$ and $\mathcal{A}_j = \mathcal{A}(D^{-1}D_j \otimes I)$.

(iii) $\widehat{\Psi}_{(p,1)}(\mathcal{A}) = \sum_j \widehat{\Psi}_{(p,1)}(\mathcal{A}_j)$

Proof: It follows from the definition of $\mathcal{A}_{123}$, that $d_2^{\frac{1+p}{p}} \Psi_{(p,1)}(\mathcal{A}) = \Phi(\mathcal{A}_{123})$. The equivalence (i) $\Leftrightarrow$ (iii) follows immediately from Theorem 15, and (i) $\Leftrightarrow$ (ii) can be shown to follow from Corollary 14. **QED**

Theorem 17 *The following monotonicity inequalities hold,*

$$\widehat{\Psi}_{(p,1)}(\mathcal{A}_{23}) \leq \widehat{\Psi}_{(p,1)}(\mathcal{A}_{123}), \quad p \in (0,2) \tag{66a}$$
$$\Psi_{(p,1)}(\mathcal{A}_{23}) \geq \Psi_{(p,1)}(\mathcal{A}_{123}), \quad p \in (0,1) \tag{66b}$$
$$\Psi_{(p,1)}(\mathcal{A}_{23}) \leq \Psi_{(p,1)}(\mathcal{A}_{123}), \quad p \in [1,2) \tag{66c}$$

Moreover, equality holds if and only if the conditions of Corollary 13 are satisfied.

Proof: It suffices to give the proof for $\widehat{\Psi}$ since the other inequalities follow immediately. The argument is similar to that for Theorem 7. Let W_n denote the generalized Pauli matrices of Section 2.3, but now let $\mathcal{W}_n = W_n \otimes I_{23}$. Then the convexity of $\widehat{\Psi}_{(p,1)}(\mathcal{A}_{23})$ implies

$$\begin{aligned}
\widehat{\Psi}_{(p,1)}(\mathcal{A}_{23}) &= \tfrac{1}{d_1}\widehat{\Psi}_{(p,1)}(I_1 \otimes \mathcal{A}_{23}) \\
&= \tfrac{1}{d_1}\widehat{\Psi}_{(p,1)}\big(\tfrac{1}{d_1}\textstyle\sum_n \mathcal{W}_n \mathcal{A}_{123} \mathcal{W}_n\big) \\
&\leq \tfrac{1}{d_1^2}\displaystyle\sum_n \widehat{\Psi}_{(p,1)}(\mathcal{W}_n \mathcal{A}_{123} \mathcal{W}_n) \;=\; \widehat{\Psi}_{(p,1)}(\mathcal{A}_{123})
\end{aligned}$$

where we used the invariance of $\widehat{\Psi}$ under unitaries of the form $U_1 \otimes I_{23}$. In the case $p = 1$, it follows from (60) that $\widehat{\Psi}_{(1,1)}(\mathcal{A}_{23}) \leq \widehat{\Psi}(1,1)(\mathcal{A}_{123})$ becomes

$$S(\mathcal{A}_2) - S(\mathcal{A}_{23}) \leq S(\mathcal{A}_{12}) - S(\mathcal{A}_{123}) \tag{67}$$

which is SSA. Because the equality conditions in Theorem 16 are independent of p, they are identical to those for SSA, which are given in Corollary 13. **QED**

The Carlen-Lieb triple Minkowski inequality for the case $q = 1$ is an immediate corollary of Theorem 17. Observe that

$$\mathrm{Tr}_3\,\mathrm{Tr}_1\big(\mathrm{Tr}_2\mathcal{A}_{123}^p\big)^{1/p} = \Psi_{(p,1)}(\mathcal{A}_{(13),(2)}) \tag{68a}$$
$$\mathrm{Tr}_3\left[\mathrm{Tr}_2(\mathrm{Tr}_1\,\mathcal{A}_{123})^p\right]^{1/p} = \Psi_{(p,1)}(\mathcal{A}_{32}) \tag{68b}$$

so that it follows immediately from (66c) that

$$\mathrm{Tr}_3\left[\mathrm{Tr}_2(\mathrm{Tr}_1\,\mathcal{A}_{123})^p\right]^{1/p} = \Psi_{(p,1)}(\mathcal{A}_{32}) \leq \Psi_{(p,1)}(\mathcal{A}_{132}) = \mathrm{Tr}_3\,\mathrm{Tr}_1\big(\mathrm{Tr}_2\mathcal{A}_{123}^p\big)^{1/p} \tag{69}$$

for $1 < p \leq 2$ and from (66b) that the inequality reverses for $0 < p < 1$. Moreover, the conditions for equality are again independent of p and identical to those for equality in SSA, given in Corollary 13.

5 Final remarks

It should be clear that the results in Section 2 are not restricted to $J_p(K, A, B)$. The function $g_p(x)$ given in (6) can be replaced by any operator convex function of the form $g(x) = xf(x)$ with f operator monotone on $(0, \infty)$. Moreover, if the measure $\nu(t)$ in (17) is supported on $(0, \infty)$, then the conditions for equality are identical to those in Section 3.

In particular, our results go through with g_p replaced by $\widetilde{g}_p$ and $J_p(I, A, B)$ replaced by $\widetilde{J}_p(I, A, B)$, which is well-defined for $p \in [-1, 1)$ with $\widetilde{J}_0(I, A, B) = H(B, A)$. Thus our results can be extended to all $p \in (-1, 2)$. The case $p = 2$ reduces to the convexity of $(A, X) \mapsto \operatorname{Tr} X^* A^{-1} X$ with $A > 0$ proved in [23]. One can show that equality holds if and only if $X_j = A_j T \quad \forall\, j$ with $T = A^{-1} X$.

There have been various attempts, e.g., the Renyi [34] and Tsallis [39] entropies, to generalize quantum entropy in a way that gives the usual von Neumann entropy at $p = 1$. In this paper we have considered two extensions of the conditional entropy involving an exponent $p \in (0, 2)$, namely,

- $J_p(I, A_{12}, A_1)$ which gives $\operatorname{Tr} A_{23}^p A_2^{1-p} \begin{array}{c} \leq \\ \geq \end{array} \operatorname{Tr} A_{123}^p A_{12}^{1-p} \begin{array}{c} p \in (0,1) \\ p \in (1,2) \end{array}$ and can be thought of as a pseudo-metric; and
- $\widehat{\Psi}_{(p,1)}(\mathcal{A}_{12})$ which gives $\operatorname{Tr}_2(\operatorname{Tr}_3 \mathcal{A}_{23}^p)^{1/p} \begin{array}{c} \geq \\ \leq \end{array} \operatorname{Tr}_{12}(\operatorname{Tr}_3 \mathcal{A}_{123}^p)^{1/p} \begin{array}{c} p \in (0,1) \\ p \in (1,2) \end{array}$ and can be thought of as a pseudo-norm.

These expressions are quite different for $p \neq 1$, but arise from quantities with the same convexity and monotonicity properties, as well as the same equality conditions which are independent of p. Moreover, both yield SSA at $p = 1$ and the equality conditions for $p \neq 1$ are identical to those for SSA. This independence of non-trivial equality conditions on the precise form of the function seems remarkable.

If one uses $\widetilde{g}_p$ and $\widetilde{J}_p(I, A, B)$ from (10), then the inequalities above hold with $p \in (1, 2)$ replaced by $p \in (-1, 0)$ and SSA corresponds to $p = 0$.

A Proof of the key Schwarz inequality

For completeness, we include the proof of the joint convexity of $(A, B, X) \mapsto \operatorname{Tr} X^*(L_A + tR_B)^{-1}(X)$ when $A, B > 0$ and $t > 0$. Since this function is homogeneous of degree one, it suffices to prove subadditivity. Now let

$$M_j = (L_{A_j} + tR_{B_j})^{-1/2}(X_j) - (L_{A_j} + tR_{B_j})^{1/2}(\Lambda). \tag{70}$$

Then one can verify that

$$\begin{aligned} 0 \;\le\; \sum_j \operatorname{Tr} M_j^* M_j \;&=\; \sum_j \langle M_j, M_j\rangle \\ =\; \sum_j \operatorname{Tr} X_j^*(L_{A_j} + tR_{B_j})^{-1}(X_j) &- \operatorname{Tr}\big(\textstyle\sum_j X_j^*\big)\Lambda \\ &-\operatorname{Tr}\Lambda^*\big(\textstyle\sum_j X_j\big) + \operatorname{Tr}\Lambda^*\textstyle\sum_j\big(L_{A_j} + tR_{B_j}\big)\Lambda. \end{aligned} \tag{71}$$

Next, observe that for any matrix W,

$$\sum_j \big(L_{A_j} + tR_{B_j}\big)(W) = \sum_j \big(A_j W + tWB_j\big) = L_{\sum_j A_j}(W) + tR_{\sum_j B_j}(W).$$

Therefore, inserting the choice $\Lambda = \big(L_{\sum_j A_j} + tR_{\sum_j B_j}\big)^{-1}\big(\sum_j X_j\big)$ in (71) yields

$$\operatorname{Tr}\big(\textstyle\sum_j X_j\big)^* \dfrac{1}{L_{\sum_j A_j} + tR_{\sum_j B_j}}\big(\textstyle\sum_j X_j\big) \le \textstyle\sum_j \operatorname{Tr} X_j^* \dfrac{1}{L_{A_j} + tR_{B_j}}(X_j). \tag{72}$$

for any $t \ge 0$. **QED**

References

[1] N. I. Akheizer and I.M. Glazman, *Theory of Operators in Hilbert Space*. Vol. II (Frederik Ungar Publishing, NY, 1963).

[2] A. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs, **191** (American Mathematical Society and Oxford University Press, 2000).

[3] T. Ando, *Topics on Operator Inequalities* Sapporo Lecture Notes (1978).

[4] T. Ando, "Concavity of certain maps on positive definite matrices and applications to Hadamard products", *Lin, Alg. Appl.* **26**, 203-241 (1979)

[5] H. Araki, "Relative Entropy of States of von Neumann Algebras" *Publ RIMS Kyoto Univ.* **9**, 809–833 (1976).

[6] T. Bekjan, "On joint convexity of trace functions" *Lin, Alg. Appl.* **390**, 321–327 (2004).

[7] E. Carlen and E. Lieb, "A Minkowski type trace inequality and strong subadditivity of quantum entropy" *Amer. Math. Soc. Translations* **189**(2), 59–62 (1999). Reprinted in [20].

[8] Eric A. Carlen, Elliott H. Lieb "A Minkowski Type Trace Inequality and Strong Subadditivity of Quantum Entropy II: Convexity and Concavity" *Lett. Math. Phys.* **83**, 107–126 (2008). arXiv:0710.4167

[9] I. Devetak and J. Yard, "Exact Cost of Redistributing Multipartite Quantum States" *Phys. Rev. Lett.* **100**, 230501 (2008).

[10] W. F. Donoghue Jr., *Monotone Matrix Functions and Analytic Continuation* (Springer, 1974).

[11] E. G. Effros, "A Matrix Convexity Approach to Some Celebrated Quantum Inequalities" *Proc. Natl. Acad. Sci.* **106**, 1006–1008 (2009). arXiv:0802.1234

[12] H. Epstein, "Remarks on Two Theorems of E. Lieb" *Commun. Math. Phys.* **31**, 317–325 (1973).

[13] P. Hayden, R. Jozsa, D. Petz, A. Winter "Structure of states which satisfy strong subadditivity of quantum entropy with equality" *Commun. Math. Phys.* **246**, 359–374, (2004)4 arXiv:quant-ph/0304007

[14] A. Jenčová and D. Petz "Sufficiency in quantum statistical inference" *Commun. Math. Phys.* **263**, 259276 (2006). arXiv:math-ph/0412093

[15] A. Jenčová and D. Petz "Sufficiency in quantum statistical inference. A survey with examples" *J. Infinite Dimensional Analysis and Quantum Probability*, **9** 331-352. (2006). arXiv:quant-ph/0604091

[16] O. Klein, "Zur quantenmechanischen begründung der zweiten hauptsatzes der wäremlehre" *Z. Phys.* **72**, 767-775 (1931).

[17] A. Lesniewski and M.B. Ruskai, "Relative Entropy and Monotone Riemannian Metrics on Non-Commutative Probability Space" *J. Math. Phys.* **40**, 5702-5724 (1999).

[18] S. Luo, N. Li and X. Cao “Relation between ‘no broadcasting’ for noncommuting states and ‘no local broadcasting’ for quantum correlations” *Phys. Rev. A* **79**, 054305 (2009).

[19] E. H. Lieb, “Convex trace functions and the Wigner-Yanase-Dyson conjecture”, *Adv. Math.* **11**, 267–288 (1973). Reprinted in [20].

[20] *Inequalities: Selecta of E. Lieb*, M. Loss and M.B. Ruskai, eds (Springer, 2002)

[21] E. H. Lieb and M.B. Ruskai, “ A Fundamental Property of the Quantum-Mechanical Entropy, *Phys. Rev. Lett.* **30**, 434–436 (1973). Reprinted in [20].

[22] E. H. Lieb and M.B. Ruskai, “Proof of the Strong Subadditivity of Quantum Mechanical Entropy” *J. Math. Phys.* **14**, 1938–1941 (1973). Reprinted in [20].

[23] E. H. Lieb and M.B. Ruskai “Some Operator Inequalities of the Schwarz Type” *Adv. Math.* **12**, 269–273 (1974). Reprinted in [20].

[24] E. H. Lieb and W. Thirring, “Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities”, in *Studies in Mathematical Physics*, E. Lieb, B. Simon, A. Wightman eds., pp. 269–303 (Princeton University Press, 1976). Reprinted in [20].

[25] G. Lindblad, “Expectations and Entropy Inequalities” *Commun. Math. Phys.* **39**, 111–119 (1974).

[26] K. Löwner, “Über monotone Matrix Funktionen” *Math. Z.* **38**, 177–216 (1934).

[27] M. Mosonyi and D. Petz, Structure of sufficient quantum coarse-grainings, *Lett. Math. Phys.* **68**(2004), 19–30.

[28] H. Narnhofer and W. Thirring, “From Relative Entropy to Entropy” *Fizika* **17**, 257–265 (1986 ?).

[29] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993, 2nd edition 2004.

[30] D. Petz, “Quasi-Entropies for Finite Quantum Systems” *Rep. Math. Phys.* **23**, 57–65 (1986).

[31] D. Petz, “Sufficiency of channels over von Neumann algebras”, *Quart. J. Math.* **39**, 907–1008 (1988), .

[32] D. Petz, “Sufficient Subalgebras and the Relative Entropy of States of a von Neumann Algebra” *Commun. Math. Phys.* **105**, 123–131 (1986).

[33] D. Petz, "Monotone Metrics on Matrix Spaces", *Lin. Alg. Appl.* **244**, 81–96 (1996).

[34] A. Rényi, "On measures of entropy and information" pp. 547–561 in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, Vol. I (Univ. California Press, Berkeley, 1961).

[35] M. B. Ruskai, "Inequalities for Quantum Entropy: A Review with Conditions for Equality" *J. Math. Phys.* **43**, 4358–4375 (2002); erratum **46**, 019901 (2005). (quant-ph/0205064).

[36] M. B. Ruskai "Another Short and Elementary Proof of Strong Subadditivity of Quantum Entropy" *Reports Math. Physics* **60**, 1–12 (2007). arXiv:quant-ph/0604206

[37] D. Ruelle, *Statistical Mechanics* (Benjamin, 1969).

[38] B. Simon, *The Statistical Mechanics of Lattice Gases* (Princeton Univ. Press, 1993).

[39] C. Tsallis, " Possible generalization of Boltzmann-Gibbs statistics" *J. Stat. Phys.* **52**, 47–4879 (1988).

[40] A. Wehrl "General Properties of Entropy" *Rev. Mod. Phys.* **50**, 221–260 (1978).

[41] E.P. Wigner and M. M. Yanase, "Information content of distributions" *Proc. Nat. Acad. Sci.* **49**, 910–918 (1963).

[42] E.P. Wigner and M. M. Yanase, "On the positive semi-definite nature of certain matrix expressions" *Canad. J. Math.* **16**, 397–406 (1964).

A Unified Treatment of Convexity of Relative Entropy and Related Trace Functions, with Conditions for Equality

Anna Jenčová *
Mathematical Institute, Slovak Academy of Sciences,
Štefánikova 49, 814 73 Bratislava, Slovakia
jenca@mat.savba.sk

Mary Beth Ruskai †
Department of Mathematics, Tufts University, Medford, MA 02155, USA
marybeth.ruskai@tufts.edu

July 30, 2010

Abstract

We consider a generalization of relative entropy derived from the Wigner-Yanase-Dyson entropy and give a simple, self-contained proof that it is convex. Moreover, special cases yield the joint convexity of relative entropy, and for $\operatorname{Tr} K^*A^pKB^{1-p}$ Lieb's joint concavity in (A,B) for $0 < p < 1$ and Ando's joint convexity for $1 < p \leq 2$. This approach allows us to obtain conditions for equality in these cases, as well as conditions for equality in a number of inequalities which follow from them. These include the monotonicity under partial traces, and some Minkowski type matrix inequalities proved by Carlen and Lieb for $\operatorname{Tr}_1(\operatorname{Tr}_2 A_{12}^p)^{1/p}$. In all cases the equality conditions are independent of p; for extensions to three spaces they are identical to the conditions for equality in the strong subadditivity of relative entropy.

*Supported by the grants VEGA 2/0032/09, APVV- 0071-06, Center of Excellence SAS - Quantum Technologies and ERDF OP R&D Project CE QUTE ITMS 26240120009
†Partially supported by National Science Foundation under Grant DMS-0604900.

1 Introduction

1.1 Background

For matrices $A_{12} > 0$ acting on a tensor product of two Hilbert spaces, Carlen and Lieb [7, 8] considered the trace function $\left[\mathrm{Tr}_1(\mathrm{Tr}_2 A_{12}^p)^{q/p}\right]^{1/q}$ and proved that it is concave when $0 \le p \le q \le 1$ and convex when $1 \le q$ and $1 \le p \le 2$. They showed that this implies that these functions and the norms they generate satisfy Minkowski type inequalities, including a natural generalization to matrices A_{123} acting on a tensor product of three Hilbert spaces. They also raised the question of the conditions for equality in their inequalities. When $q = 1$, we show that this can be treated using methods developed to treat equality in the strong subadditivity of quantum entropy. Moreover, we obtain conditions for equality in a large class of related convexity inequalities, show that they are independent of p in the range $0 < p < 2$, and show that for inequalities involving A_{123} they are identical to the equality conditions for strong subadditivity (SSA) of quantum entropy given in [13]

These equality conditions are non-trivial and have found many applications in quantum information theory. For example, they play an important role in some recent "no broadcasting" results; see [19] and references therein. They also play a key role in Devetak and Yard's [9] "quantum state redistribution" protocol which gives an operational interpretation to the quantum conditional mutual information.

Our approach to proving joint convexity of relative entropy is motivated by Araki's relative modular operator [5], introduced to generalize relative entropy to more general situations including type III von Neumann algebras. It was subsequently used by Narnhofer and Thirring [29] to give a new proof of SSA. The argument given here is similar to that in [18, 31, 37]; however, the unified treatment for $0 < p < 2$ leading to equality conditions, is new. Moreover, a dual treatment can be given for $-1 < p < 1$ allowing extension to the full range $(-1, 2)$.

Wigner and Yanase [42, 43] introduced the notion of skew information of a density matrix γ with respect to a self-adjoint observable K,

$$-\,\mathrm{Tr}\,\tfrac{1}{2}[K,\gamma^p]\,[K,\gamma^{1-p}] \tag{1}$$

for $p = \frac{1}{2}$ and Dyson suggested extending this to $p \in (0,1)$. Wigner and Yanase [43] proved that (1) is convex in γ for $p = \frac{1}{2}$ and, in his seminal paper [20] on convex trace functions, Lieb proved joint concavity for $p \in (0,1)$ for the more general function

$$(A,B) \mapsto \mathrm{Tr}\,K^* A^p K B^{1-p} \tag{2}$$

for K fixed and $A, B > 0$ positive semi-definite. This implies convexity of (1) and was a key step in the original proof [23] of the strong subadditivity (SSA) inequality

of quantum entropy. Moreover, it leads to a proof of joint convexity of relative entropy[1] as well. It is less well known that Ando [3, 4] gave another proof which also showed that for $1 \le p \le 2$, the function (2) is jointly convex in A, B. The case $p = 2$ was considered earlier by Lieb and Ruskai [24]. We modify what one might describe as Lieb's extension of the Wigner-Yanase-Dyson (WYD) entropy to a type of relative entropy in a way that allows a unified treatment of the convexity and concavity of $\operatorname{Tr} K^* A^p K B^{1-p}$ in the range $p \in (0, 2]$ and includes the usual relative entropy as a special case. Our modification retains a linear term, even for $A \ne B$. Although this might seem unnecessary for convexity and concavity questions, it is crucial to a unified treatment.

Lieb also considered $\operatorname{Tr} K^* A^p K B^q$ with $p, q > 0$ and $0 \le p + q \le 1$ and Ando considered $1 < q \le p \le 2$. In Section 2.2, we extend our results to this situation. However, we also show that for $q \ne 1-p$, equality holds only under trivial conditions. Therefore, we concentrate on the case $q = 1 - p$.

Next, we introduce our notation and conventions. In Section 2, we first describe our generalization of relative entropy and prove its convexity; then consider the extension to $q \ne 1-p$ mentioned above; and finally prove monotonicity under partial traces including a generalization of strong subadditivity to $p \ne 1$. In Section 3, we consider several formulations of equality conditions. In Section 4, we show how to use these results to obtain equality conditions in the results of Lieb and Carlen [7, 8]. For completeness, we include an appendix which contains the proof of a basic convexity result from [37] that is key to our results.

1.2 Notation and conventions

We introduce two linear maps on the space M_d of $d \times d$ matrices. Left multiplication by A is denoted L_A and defined as $L_A(X) = AX$; right multiplication by B is denoted R_B and defined as $R_B(X) = XR$. These maps are associated with the relative modular operator $\Delta_{AB} = L_A R_B^{-1}$ introduced by Araki in a far more general context. They have the following properties:

a) The operators L_A and R_B commute since

$$L_A[R_B(X)] = AXB = R_B[L_A(X)] \tag{3}$$

even when A and B do not commute.

[1] In [23] only concavity of the conditional entropy was proved explicitly, but the same argument [36, Section V.B] yields joint convexity of the relative entropy. Independently, Lindblad [26] observed that this follows directly from (2) by differentiating at $p = 1$.

b) L_A and R_A are invertible if and only if A is non-singular, in which case $L_A^{-1} = L_{A^{-1}}$ and $R_A^{-1} = R_{A^{-1}}$.

c) When A is self-adjoint, L_A and R_A are both self-adjoint with respect to the Hilbert Schmidt inner product $\langle A, B\rangle = \text{Tr}\, A^*B$.

d) When $A \geq 0$, the operators L_A and R_A are positive semi-definite, i.e.,

$$\begin{aligned} \text{Tr}\, X^* L_A(X) = \text{Tr}\, X^*AX \geq 0 \qquad \text{and} \\ \text{Tr}\, X^* R_A(X) = \text{Tr}\, X^*XA = \text{Tr}\, XAX^* \geq 0. \end{aligned}$$

e) When $A > 0$, then $(L_A)^p = L_{A^p}$ and $(R_A)^p = R_{A^p}$ for all $p \geq 0$. If A is also non-singular, this extends to all $p \in \mathbf{R}$. More generally $f(L_A) = L_{f(A)}$ for $f : (0, \infty) \mapsto \mathbf{R}$.

To see why (e) holds, it suffices to observe that $A > 0$ implies L_A and R_A are linear operators for which $f(A)$ can be defined by the spectral theorem for any function f with domain in $(0, \infty)$. It is easy to verify that $A|\phi_j\rangle = \alpha_j|\phi_j\rangle$ implies $L_A|\phi_j\rangle\langle\phi_k| = \alpha_j|\phi_j\rangle\langle\phi_k|$ for $k = 1 \ldots d$ so that the spectral decomposition of A induces one on L_A with degeneracy d and $f(L_A)|\phi_j\rangle\langle\phi_k| = f(\alpha_k)|\phi_j\rangle\langle\phi_k|$. For R_B a similar argument goes through starting with left eigenvectors of B i.e., $\langle\phi_j|B = \beta_j\langle\phi_j|$.

If a function is homogeneous of degree 1, then convexity is equivalent to subadditivity. Thus, if $F(\lambda A) = \lambda F(A)$, then F is convex if and only if $F(A) \leq \sum_j F(A_j)$ with $A = \sum_j A_j$. We will use this equivalence without further ado.

For B positive semi-definite, we denote the projection onto $(\ker B)^\perp$ by $P_{(\ker B)^\perp}$. We will encounter expressions involving commuting positive semi-definite matrices A, D with $\ker D \subseteq \ker A$. We will simply write AD^{-1} for

$$\lim_{\epsilon \to 0} \sqrt{A}(D + \epsilon I)^{-1}\sqrt{A} = AD^{-1}P_{(\ker D)^\perp} = AD^{-1}P_{(\ker A)^\perp} \tag{4}$$

with D^{-1} the generalized inverse.

2 WYD entropy revisited and extended

2.1 Generalization of relative entropy

We now introduce the family of functions

$$g_p(x) = \begin{cases} \frac{1}{p(1-p)}(x - x^p) & p \neq 1 \\ x \log x & p = 1 \end{cases}. \tag{5}$$

which are well-defined for $x > 0$ and $p \neq 0$. We will consider $p \in (0,2]$ although it would suffice to consider $p \in [\frac{1}{2}, 2]$. For A, B strictly positive we define

$$\begin{aligned} J_p(K,A,B) &\equiv \operatorname{Tr} \sqrt{B} K^* \, g_p\big(L_A R_B^{-1}\big)(K\sqrt{B}) && (6) \\ &= \begin{cases} \frac{1}{p(1-p)}\big(\operatorname{Tr} K^*AK - \operatorname{Tr} K^*A^pKB^{1-p}\big) & p \in (0,1)\cup(1,2) \\ \operatorname{Tr} KK^*A\log A - \operatorname{Tr} K^*AK\log B & p=1 \\ -\frac{1}{2}\big(\operatorname{Tr} K^*AK - \operatorname{Tr} AKB^{-1}K^*A\big) & p=2 \end{cases} && (7) \end{aligned}$$

When $p = 1$ and $K = I$, (6) reduces to the usual relative entropy, i.e.,

$$J_1(I,A,B) = H(A,B) = \operatorname{Tr} A(\log A - \log B) \tag{8}$$

For $p \neq 1$, the function $J_p(K,A,B)$ differs from that considered by Lieb [20] and Ando [3, 4] by the seemingly irrelevant linear term $\operatorname{Tr} K^*AK$ and the factor $\frac{1}{p(1-p)}$. However, this minor difference allows us to give a unified treatment of $p \in (0,2]$ because of the extension by continuity to $p = 1$ and the sign change there.

One might expect to associate the exchange $A \leftrightarrow B$ with the symmetry $p \leftrightarrow (1-p)$ around $p = \frac{1}{2}$. However, there are several subtleties due to the linear term, the exchange $K \leftrightarrow K^*$, and the case $p = 1$. Therefore, we use instead the observation that

$$\begin{aligned} J_p(K^*,B,A) &= \operatorname{Tr} \sqrt{A}\, K\, g_p\big(L_B R_A^{-1}\big)(K^*\sqrt{A}) \\ &= \operatorname{Tr} \sqrt{B} K^* \, \widetilde{g}_{1-p}\big(L_A R_B^{-1}\big)(K\sqrt{B}) \\ &= \widetilde{J}_{1-p}(K,A,B) && (9) \end{aligned}$$

where, for $-1 \leq p < 1$, we define

$$\widetilde{g}_p(x) = x g_{1-p}(x^{-1}) = \begin{cases} \frac{1}{p(1-p)}(1-x^p) & p \neq 0 \\ -\log x & p = 0 \end{cases} \tag{10}$$

and $\widetilde{J}_p(K,A,B) = \operatorname{Tr} \sqrt{B} K^* \, \widetilde{g}_p\big(L_A R_B^{-1}\big)(K\sqrt{B})$.

The functions $J_p(K,A,B)$ and $\widetilde{J}_p(K,A,B)$ have been considered before, usually with $K = I$, in the context of information geometry ([2, Section 7.2] and references therein) and by Petz [31] who used the term "quasi-entropy". What is novel here is that we present a simple unified proof of joint convexity in A, B that easily yields equality conditions, shows that they are independent of p, and can be extended to other functions.

The special case $J_p(I,A,I)$ is equivalent[2] to the Tsallis [40] entropy. When $K = K^*$, the relation

$$J_p(K,A,A) = -\frac{1}{2p(1-p)} \operatorname{Tr}\, [K,A^p][K,A^{1-p}] \tag{11}$$

[2]This was pointed out by Karol Zyczkowski

yields the original WYD information (up to a constant) and extends it to the range $(0,2]$. Morevoer, $K = K^*$ implies that $J_p(K,A,A) = \widetilde{J}_{1-p}(K,A,A)$. Although neither $g_p(w)$ nor $\widetilde{g}_{1-p}(w)$ is positive, their average[3]

$G_p(w) \equiv \frac{1}{2}[g_p(w) + w g_p(w^{-1})] \geq 0$ on $(0,\infty)$. Therefore, when $K = K^*$,

$$J_p(K,A,A) = \mathrm{Tr}\,(K\sqrt{A})^* G_p(L_A R_A^{-1})(K\sqrt{A}) \geq 0 \tag{12}$$

The function $J_p(I,A,B)$ is a more appealing generalization of relative entropy than $\mathrm{Tr}\,A^p B^{1-p}$ because of Proposition 1, which one can consider to be a generalization of Klein's inequality [17]. It allows one to use $J_p(I,A,B)$ as a pseudo-metric, as is commonly done with the relative entropy.

Proposition 1 *When U is unitary and $A, B > 0$ with $\mathrm{Tr}\,A = \mathrm{Tr}\,B = 1$, then $J_p(U,A,B) \geq 0$ with equality if and only if $A = U^*BU$.*

Proof: When U is unitary,

$$J_p(U,A,B) = J_p(I, U^*AU, B) = J_p(I, A, UBU^*). \tag{13}$$

Therefore, it suffices to consider the case $U = I$. For $p \in (0,1)$ Hölders inequality implies $\mathrm{Tr}\,A^p B^{1-p} \leq (\mathrm{Tr}\,A)^p (\mathrm{Tr}\,B)^{1-p} = 1$ with equality if and and only $A = B$. It immediately follows that

$$J_p(I,A,B) \geq \tfrac{1}{p(1-p)}\big(\mathrm{Tr}\,A - 1\big) = 0 \qquad \text{and} \qquad J_p(I,A,B) = 0 \Leftrightarrow A = B. \tag{14}$$

For $p = 1$, the result is well-known [38, Section 2.5.2] and originally due to O. Klein [17]. For $p \in (1,2)$ we write $p = 1 + r$ and again use Hölder's inequality

$$\begin{aligned} 1 &= \mathrm{Tr}\,A = \mathrm{Tr}\,B^{-\frac{r}{2(r+1)}} A B^{-\frac{r}{2(r+1)}} B^{\frac{r}{r+1}} \\ &\leq \Big[\mathrm{Tr}\,\big(B^{-\frac{r}{2(r+1)}} A B^{-\frac{r}{2(r+1)}}\big)^{1+r}\Big]^{\frac{1}{1+r}} (\mathrm{Tr}\,B)^{\frac{r}{1+r}} \\ &\leq \big[\mathrm{Tr}\,B^{-\frac{1}{2}} A^{1+r} B^{-\frac{1}{2}}\big]^{\frac{1}{1+r}} \big(\mathrm{Tr}\,A^{1+r} B^{-r}\big)^{\frac{1}{1+r}} \end{aligned} \tag{15}$$

where we used $\mathrm{Tr}\,B = 1$ and the second inequality follows from a classic result of Lieb-Thirring [25, Appendix B, Theorem 9]. **QED**

Because the denominator $p(1-p)$ changes sign at $p = 0$ and $p = 1$, both g_p and $\widetilde{g}_p$ are convex. In fact, they satisfy the much stronger condition of operator convexity for $p \in (0,2]$ and $p \in [-1,1)$ respectively. Since $g(0) = 0$ and

$$\frac{g_p(x)}{x} = \begin{cases} \frac{1}{p(1-p)}(1 - x^{p-1}) & p \neq 1 \\ \log x & p = 1 \end{cases}, \tag{16}$$

[3]The definition of $\widetilde{g}_p$ in (10) differs from that in [18] by the exchange $\widetilde{g}_p \leftrightarrow \widetilde{g}_{1-p}$ so that in [18] $G(w) = \frac{1}{2}[g(w) + \widetilde{g}(w)]$ for any g. In the convention used here, $G_p(w) = \frac{1}{2}[g_p(w) + \widetilde{g}_{1-p}(w)]$.

it follows that $g_p(x)/x$ is operator monotone [3, 10, 27], for $p \in (0,2]$, i.e., g_p can be analytically continued to the upper half plane, which it maps into itself. By applying Nevanlinna's theorem [1, Section 59, Theorem 2] to $g_p(x)/x$, one finds that $g_p(x)$ has an integral representation of the form

$$\begin{aligned} g_p(x) &= ax + \int_0^\infty \frac{x^2 t - x}{x+t}\, d\nu(t) \\ &= ax + \int_0^\infty \Big[\frac{x^2}{x+t} - \frac{1}{t} + \frac{1}{x+t}\Big]\, t\, d\nu(t) \end{aligned} \tag{17}$$

with $\nu(t) \geq 0$. Integral representations are not unique, and making a suitable change of variable in the classic formula

$$\int_0^\infty \frac{x^{p-1}}{x+1} = \frac{\pi}{\sin p\pi} \equiv \frac{1}{c_p} \qquad\qquad p \in (0,1) \tag{18}$$

allows us to give the following explicit representations

$$g_p(x) = \begin{cases} \frac{1}{p(1-p)}\Big[x + c_p \int_0^\infty \big(\frac{t}{x+t} - 1\big) t^{p-1} dt\Big] & p \in (0,1) \\ \\ \int_0^\infty \big(\frac{x^2}{x+t} - 1 + \frac{t}{x+t}\big)\frac{1}{1+t} dt & p = 1 \\ \\ \frac{1}{p(1-p)}\Big[x - c_{p-1}\int_0^\infty \frac{x^2}{x+t} t^{p-2} dt\Big] & p \in (1,2) \\ \\ \frac{1}{2}(-x + x^2) & p = 2 \end{cases} \tag{19}$$

Note that for $p \in (0,2)$ the integrand is supported on $(0,\infty)$. This plays a key role in the equality conditions; therefore, we will henceforth concentrate on $p \in (0,2)$.

Theorem 2 *The function $J_p(K,A,B)$ defined in (6) is jointly convex in A, B.*

Proof: It follows from (17) that

$$\begin{aligned} J_p(K,A,B) &= a \operatorname{Tr} K^* A K \\ &+ \int_0^\infty \Big[\operatorname{Tr} K^* A \frac{1}{L_A + tR_B}(AK) - \frac{\operatorname{Tr} KBK^*}{t} + \operatorname{Tr} BK^* \frac{1}{L_A + tR_B}(KB)\Big] t\, \nu(t)\, dt \end{aligned} \tag{20}$$

The joint convexity then follows immediately from that of the map $(X,A,B) \mapsto \operatorname{Tr} X^* \frac{1}{L_A + tR_B}(X)$ which was proved in [37] following the strategy in [24]. The proof is also given in the Appendix. **QED**

For other approaches see Petz [30, 31], Effros [11], The advantage to the argument used here is that it immediately implies that equality holds in joint convexity if and only if it holds for each term in the integrand.

Corollary 3 *The relative entropy $H(A,B) = J_1(I,A,B)$ is jointly convex in A, B.*

2.2 Extensions with $r \neq 1-p$.

We now consider extensions of Theorem 2 to situations considered by Ando [4] and Lieb [20] in which B^{1-p} is replaced by B^r with $r \neq 1-p$. Our approach uses an idea from Bekjan [6] and Effros [11]. We will also show that equality holds in these extensions only under trivial conditions. For this we first need an elementary lemma, which we prove for the concave case.

Lemma 4 *Let $f(\lambda) : [0,\infty) \mapsto \mathbf{R}$ be a non-linear convex or concave operator function, let A_1, A_2 be density matrices and $A = \lambda A_1 + (1-\lambda)A_2$ with $\lambda \in (0,1)$. Then $f(A) = \lambda f(A_1) + (1-\lambda) f(A_2)$ if and only if $A_1 = A_2$.*

Proof: Since any operator concave function is analytic, non-linearity implies that f is strictly concave. If $f(A) = \lambda f(A_1) + (1-\lambda)f(A_2)$, then

$$\langle v, f(A)v\rangle = \lambda\langle v, f(A_1)v\rangle + (1-\lambda)\langle v, f(A_2)v\rangle \tag{21}$$

for any vector v. Now choose v to be a normalized eigenvector of A. Then inserting this on the left above and applying Jensen's inequality to each term on the right, one finds

$$f\big(\langle v, Av\rangle\big) \leq \lambda f\big(\langle v, A_1 v\rangle\big) + (1-\lambda) f\big(\langle v, A_2 v\rangle\big) \tag{22}$$

But this contradicts concavity unless equality holds, which implies that v is also an eigenvector of A_1 and A_2. But then the strict concavity of f also implies that $\langle v, A_1 v\rangle = \langle v, A_2 v\rangle$. Since this holds for an orthonormal basis of eigenvectors of A, A_1 and A_2, we must have $A_1 = A_2$.

Corollary 5 *The function $(A,B) \mapsto \operatorname{Tr} K^* A^p K B^r$ is jointly concave on the set of positive definite matrices when $p, r \geq 0$ and $p + r \leq 1$. Moreover, when $p + r < 1$ and K is invertible, the convexity is strict unless $B_1 = B_2$ and $A_1 = A_2$.*

Proof: It is an immediate consequence of Theorem 2 that $(A,B) \mapsto \operatorname{Tr} K^* A^p K B^{1-p}$ is jointly concave in A, B. Now write $\operatorname{Tr} K^* A^p K B^r = \operatorname{Tr} K^* A^p K (B^s)^{1-p}$ with $s = r/(1-p)$. First, observe that for $0 < s < 1$ the function $f(x) = x^s$ satisfies the hypotheses of Lemma 4. Therefore,

$$\big(\lambda B_1 + (1-\lambda)B_2\big)^s > \lambda B_1^s + (1-\lambda)B_2^s \tag{23}$$

with $0 < \lambda < 1$ and $B_1 \neq B_2$. The operator monotonicity of $x \mapsto x^{1-p}$ for $0 < p < 1$ then implies

$$\big(\lambda B_1 + (1-\lambda)B_2\big)^r > \big(\lambda B_1^s + (1-\lambda)B_2^s\big)^{1-p}, \tag{24}$$

and the joint concavity of $\operatorname{Tr} K^* A^p K B^{1-p}$ implies

$$\begin{aligned}\operatorname{Tr} K^* A^p K (B^s)^{1-p} &\geq \operatorname{Tr} K^* \big(\lambda A_1 + (1-\lambda) A_2\big)^p K \big(\lambda B_1^s + (1-\lambda) B_2^s\big)^{1-p} \\ &\geq \lambda \operatorname{Tr} K^* A_1^p K B_1^{s(1-p)} + (1-\lambda) \operatorname{Tr} K^* A_2^p K B_2^{s(1-p)}\end{aligned} \tag{25}$$

where $A = \lambda A_1 + (1-\lambda) A_2$, $B = \lambda B_1 + (1-\lambda) B_2$, which is precisely the joint concavity of $\operatorname{Tr} K^* A^p K B^r$. Moreover, equality in joint concavity implies equality in (25) and, since $K^* A^p K$ is strictly positive, this implies equality in (23). Therefore, equality in (25) gives a contradiction unless $B_1 = B_2$. In that case, the joint concavity reduces to concavity in A for which, by a similar argument, equality holds if and only if $A_1 = A_2$. **QED**

Corollary 6 *The function $(A, B) \mapsto \operatorname{Tr} K^* A^p K B^{1-r}$ is jointly convex on the set of positive definite matrices when $1 < r \leq p \leq 2$. Moreover, when $r < p$ and K is invertible, the convexity is strict unless $B_1 = B_2$ and $A_1 = A_2$.*

Proof: The argument is similar to that for Corollary 5. Write $\operatorname{Tr} K^* A^p K B^{1-r} = \operatorname{Tr} K^* A^p K (B^s)^{1-p}$ with $s = \frac{1-r}{1-p}$. Since $s \in (0, 1)$ and $1 - p \in (-1, 0)$ when $1 < r < p < 2$, it follows that x^s is operator concave and x^{1-p} is operator monotone decreasing. **QED**

2.3 Monotonicity under partial traces

Let X and Z denote the generalized Pauli operators whose action on the standard basis is $X|e_k\rangle = |e_{k+1}\rangle$ (with subscript addition mod d) and $Z|e_k\rangle = e^{i2\pi k/d}|e_k\rangle$. It is well known and easy to verify that $\frac{1}{d}\sum_k Z^k A Z^{-k}$ is the projection of a matrix onto its diagonal. If D is a diagonal matrix, then $\sum_k X^k D X^{-k} = (\operatorname{Tr} D) I$. Now let $\{W_n\}_{n=1,2\ldots d^2}$ denote some ordering of the generalized Pauli operators , e.g., $W_{j+k(d-1)} = X^j Z^k$ with $j, k = 1, 2 \ldots d$. Then $\frac{1}{d}\sum_n W_n A W_n^* = (\operatorname{Tr} A) I$ and

$$\tfrac{1}{d}\sum_n (W_n \otimes I_2)\, A_{12}\, (W_n \otimes I_2)^* = I_1 \otimes (\operatorname{Tr}_1 A) = I_1 \otimes A_2 \tag{26}$$

Using the fact that replacing W_n by $U W_n U^*$ with U unitary, simply corresponds to a change of basis which does not affect (26) and then multiplying both sides by $U^* \otimes I_2$ on the left and $U \otimes I_2$ on the right gives the equivalent expression

$$\tfrac{1}{d}\sum_n (W_n U^* \otimes I_2)\, A_{12}\, (W_n U^* \otimes I_2)^* = I_1 \otimes A_2 \tag{27}$$

Combining this with joint convexity yields a slight generalization of the well-known monotonicity of $J_p(K, A, B)$ under partial traces (MPT), first proved by Lieb in [20] for the case $K_{12} = I_1 \otimes K_2$ when $p \in (0, 1)$.

Theorem 7 *Let J_p be as in (7), A_{12}, B_{12} strictly positive in $M_{d_1} \otimes M_{d_2}$ and $K_{12} = V_1 \otimes K_2$ with V_1 unitary in M_{d_1}. Then*

$$J_p(K_2, A_2, B_2) \leq J_p(K_{12}, A_{12}, B_{12}) \tag{28}$$

Proof: Writing $\mathcal{W}_n$ for $W_n \otimes I_2$ and using (27) gives

$$\begin{aligned}
J_p(K_2, A_2, B_2) &= \tfrac{1}{d_1} J_p(I_1 \otimes K_2, I_1 \otimes A_2, I_1 \otimes B_2) \\
&= \tfrac{1}{d_1} J_p\Big(I_1 \otimes K_2,\ \tfrac{1}{d_1} \textstyle\sum_n \mathcal{W}_n (V_1^* \otimes I_2) A_{12} (V_1 \otimes I_2) \mathcal{W}_n^*,\ \tfrac{1}{d_1} \textstyle\sum_n \mathcal{W}_n B_{12} \mathcal{W}_n^*\Big) \\
&\leq \tfrac{1}{d_1^2} \sum_n J_p\big(I_1 \otimes K_2, \mathcal{W}_n (V_1^* \otimes I_2) A_{12} (V_1 \otimes I_2) \mathcal{W}_n^*, \mathcal{W}_n B_{12} \mathcal{W}_n^*\big) \\
&= J_p(V_1 \otimes K_2, A_{12}, B_{12})
\end{aligned}$$

where the final equality follows from the unitary invariance of the trace. **QED**

Because $\mathrm{Tr}_{\,12}(V_1 \otimes K_2) A_{12} (V_1 \otimes K_2)^* = \mathrm{Tr}_{\,2} K_2 A_2 K_2^*$, (28) is equivalent to

$$\mathrm{Tr}\, K_2^* A_2^p K_2 B_2^{1-p} - \mathrm{Tr}\,(V_1 \otimes K_2)^* A_{12}^p (V_1 \otimes K_2) B_{12}^{1-p} \begin{cases} \geq 0 & p \in (0,1) \\ \leq 0 & p \in (1,2) \end{cases}. \tag{29}$$

We can obtain a weak reversal of this for $p \in (0,1)$. The argument in the Appendix shows that for any p and fixed $A, B \geq 0$ both $\mathrm{Tr}\, K^* A^p K B^{1-p}$ and $\mathrm{Tr}\, K^* A K$ are convex in K. This was observed earlier by Lieb [20] and also follows from the results in [24]. One can then apply the argument above in the special case $A_{12} = I_1 \otimes A_2, B_{12} = I_1 \otimes B_2$ to conclude that

$$\begin{aligned}
\mathrm{Tr}\, K_2^* A_2^p K_2 B_2^{1-p} &\leq \tfrac{1}{d_1} \mathrm{Tr}\, K_{12}^* (I_1 \otimes A_2)^p K_{12} (I_1 \otimes B_2)^{1-p} && (30) \\
&\leq \mathrm{Tr}\, K_{12}^* (I_1 \otimes A_2)^p K_{12} (I_1 \otimes B_2)^{1-p} && (31)
\end{aligned}$$

independent of whether $p < 1$ or $p > 1$. However, because the term $\mathrm{Tr}\, K^* A K$ is convex rather than linear in K, (30) does not allow us to draw any conclusions about the monotonicity of $J_p(K_{12}, I_1 \otimes A_2, I_1 \otimes B_2)$.

To prove Theorem 7 we showed that joint convexity implies monotonicity; the reverse implication also holds. Let $A_1, \ldots, A_m$, $B_1, \ldots, B_m$ be positive definite matrices in M_d, $A = \sum_j A_j$, $B = \sum_j B_j$, and put

$$\widetilde{A}_{12} = \sum_j |e_j\rangle\langle e_j| \otimes A_j, \quad \widetilde{B}_{12} = \sum_j |e_j\rangle\langle e_j| \otimes B_j, \tag{32}$$

for $e_1, \ldots, e_m$ the standard basis of $\mathbf{C}^m$. Then $\widetilde{A}_{12}$ and $\widetilde{B}_{12}$ are block diagonal, and $\widetilde{A}_2 = \mathrm{Tr}_1 \widetilde{A}_{12} = \sum_k A_k = A$ and similarly for B. Then if monotonicity under partial

traces holds, one can conclude that

$$\begin{aligned} J_p(K,A,B) &= J_p(K,\widetilde{A}_2,\widetilde{B}_2) \\ &\le J_p(I_1\otimes K,\widetilde{A}_{12},\widetilde{B}_{12}) = \sum_j J_p(K,A_j,B_j) \end{aligned} \tag{33}$$

Thus, monotonicity under partial traces also directly implies joint convexity of J_p.

Applying (28) in the case $K=I$, and $A_{12}\mapsto A_{123}$ and $B_{12}\mapsto A_{12}\otimes I_3$ gives

$$J_p(I_{23},A_{23},A_2\otimes I_3) \le J_p(I_{123},A_{123},A_{12}\otimes I_3) \tag{34}$$

When $p=1$, it follows from (7) that

$$J_1(I_{23},A_{23},A_2\otimes I_3) = H(A_{23},A_2\otimes I_2) = -S(A_{23})+S(A_2)$$

where $S(A)=-\mathrm{Tr}\,A\log A$. Thus, (34) becomes

$$-S(A_{23})+S(A_2) \le -S(A_{123})+S(A_{12})$$

or, equivalently

$$S(A_2)+S(A_{123}) \le S(A_{12})+S(A_{23}) \tag{35}$$

which is the standard form of SSA.

3 Equality for joint convexity of $J_p(K,A,B)$.

3.1 Origin of necessary and sufficient conditions

Looking back at the proof of Theorem 2, we see that for $p\in(0,2)$, equality holds in the joint convexity of $J_p(K,A,B)$ if and only if equality holds in the joint convexity for each term in the integrand in (17). It should be clear from the argument given in the Appendix, that this requires $M_j=0$ for all j with M_j given by (70). This is easily seen to be equivalent to

$$(L_{A_j}+tR_{B_j})^{-1}(X_j) = (L_A+tR_B)^{-1}(X) \quad \text{for all } j. \tag{36}$$

with $A=\sum_j A_j$, $B=\sum_j B_j$, and $X=\sum_j X_j$ with $X_j=A_jK$ and/or $X_j=KB_j$. By writing $AK=L_A(K)$ in the former case and $KB=R_B(K)$ in the latter we obtain the conditions

$$(I+t\Delta_{A_jB_j}^{-1})^{-1}(K) = (I+t\Delta_{AB}^{-1})^{-1}(K) \quad \forall\, j \quad \forall t>0 \tag{37a}$$

$$(\Delta_{A_jB_j}+tI)^{-1}(K) = (\Delta_{AB}+tI)^{-1}(K) \quad \forall\, j \quad \forall t>0 \tag{37b}$$

From the integral representations (19), one might expect it to be necessary for either or both of (37a) and (37b) to hold depending on p. In fact, either will suffice because (37a) holds if and only if (37b) holds. Because Δ_{AB} is positive definite, by analytic continuation (37b) extends from $t > 0$ to the entire complex plane, except points $-t$ on the negative real axis for which $t \in \text{spectrum}(\Delta_{AB})$. Therefore, by using the Cauchy integral formula, one finds that for any function G analytic on $\mathbf{C}\backslash(-\infty, 0]$ $G(\Delta_{A_jB_j})(K) = G(\Delta_{AB})(K)$.

Theorem 8 *For fixed K, and $A = \sum_j A_j, B = \sum_j B_j$, the following are equivalent*

a) $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ *for all* $p \in (0, 2)$.

b) $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ *for some* $p \in (0, 2)$.

c) $(\Delta_{A_jB_j} + tI)^{-1}(K) = (\Delta_{AB} + tI)^{-1}(K)$ *for all* j *and for all* $t > 0$.

d) $A_j^{it} K B_j^{-it} = A^{it} K B^{-it}$ *for all* j *and for all* $t > 0$.

e) $(\log A - \log A_j)K = K(\log B - \log B_j)$ *for all* j.

Proof: Clearly (a) $\Rightarrow$ (b). The implications (b) $\Rightarrow$ (c) $\Rightarrow$ (d), as well as (b) $\Rightarrow$ (a), follow from the discussion above. Differentiation of (d) at $t = 0$ gives (d) $\Rightarrow$ (e), and it is straightforward to verify that (e) $\Rightarrow$ (b) with $p = 1$. Moreover, (d) implies $\sum_j \text{Tr}\, K^* A_j^{it} K B_j^{1-it} = \text{Tr}\, K^* A^{it} K B^{1-it}$ for all t, which implies (a) by analytic continuation. **QED**

3.2 Sufficient subalgebras

When $K = I$, we can obtain a more useful reformulation of the equality conditions by using results about sufficient subalgebras obtained in [14, 15, 33]. Since the definition and convexity properties of $J_p(I, A, B)$ extend by continuity to positive semidefinite matrices, with $\ker B \subseteq \ker A$, we will formulate the conditions in this more general situation, using the conventions in Section 1.2.

Let $N \subseteq M_d$ be a subalgebra, then there is a trace preserving conditional expectation E_N from M_d onto N, such that $\text{Tr}\, AX = \text{Tr}\, E_N(A)X$ for all $X \in N$. In particular, if $N = M_{d_1} \otimes I \subseteq M_{d_1} \otimes M_{d_2}$, then we have $E_N(A_{12}) = \text{Tr}_2 A \otimes \frac{1}{d_2} I$.

Let $Q_1, \ldots, Q_m \in M_d^+$ and assume that $\ker Q_m \subseteq \ker Q_j$ for all j. The subalgebra N is said to be sufficient for $\{Q_1, \ldots, Q_m\}$ if there is a completely positive trace preserving map $T : N \to M_d$, such that $T(E_N(Q_j)) = Q_j$ for all $j = 1, \ldots, m$. This definition is due to Petz [33, 32] and it is a quantum generalization of the well known notion of sufficiency from classical statistics. In [33], it was shown that sufficient subalgebras can be characterized by the condition

$$H(Q_j, Q_m) = H(E_N(Q_j), E_N(Q_m)), \quad \text{for all } j$$

We combine this with the results of the previous section to obtain other useful characterizations of sufficiency.

Theorem 9 *Let $Q_1, \ldots, Q_m \in M_d^+$ be such that $\ker Q_m \subseteq \ker Q_j$ for all j. Let $N \subseteq M_d$ be a subalgebra. The following are equivalent.*

(i) N is sufficient for $\{Q_1, \ldots, Q_m\}$.

(ii) $E_N(Q_j)^{it} E_N(Q_m)^{-it} P_{(\ker Q_m)^\perp} = Q_j^{it} Q_m^{-it}$, for all j, $t \in \mathbf{R}$.

(iii) There exist $Q_{j,0} \in N^+$, and $D \in M_d^+$, such that $\ker D = \ker Q_m$, and $Q_j = Q_{j,0} D$ for $j = 1, \ldots, m$.

(iv) $J_p(I, Q_j, Q_m) = J_p(I, E_N(Q_j), E_N(Q_m))$ for all j and some $p \in (0,1)$

The proof of the conditions (i) – (iii) can be found in [14], see also [28]. The condition (iv) was proved in [15].

3.3 Equality conditions with $K = I$

Theorem 10 *Let $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ be positive semi-definite matrices with $\ker B_j \subseteq \ker A_j$, and let $A = \sum_j A_j, B = \sum_j B_j$. Then the following are equivalent.*

a) $J_p(I, A, B) = \sum_j J_p(I, A_j, B_j)$ for all $p \in (0, 2)$.

b) $J_p(I, A, B) = \sum_j J_p(I, A_j, B_j)$ for some $p \in (0, 2)$.

c) $A_j^{it} B_j^{-it} = A^{it} B^{-it} P_{(\ker B_j)^\perp}$ for all j and $t \in \mathbf{R}$

d) There are positive matrices $D_1, \ldots, D_m$, with $\ker D_j = \ker B_j$, such that $[A_j, D_j] = [B_j, D_j] = 0$, and with $D = \sum_j D_j$

$$A_j = AD^{-1}D_j, \qquad B_j = BD^{-1}D_j \tag{38}$$

Proof: As in Section 3.1, (b) implies (36) on $(\ker B_j)^\perp$, with $X_j = B_j, X = B$. This gives

$$(\Delta_{A_j B_j} + tI)^{-1}(I) = (\Delta_{AB} + tI)^{-1}(I) \qquad \text{on} \quad (\ker B_j)^\perp. \tag{39}$$

Then (c) follows from the Cauchy integral formula as in Section 3.1.

To show (c) implies (d), we will use Theorem 9. First let $N = I \otimes M_d \subseteq M_m \otimes M_d$ and let $\widetilde{A}_{12}$, $\widetilde{B}_{12}$ be the block-diagonal matrices in $M_m \otimes M_d$, defined by (32). Clearly, we have $\ker \widetilde{A}_{12} \supseteq \ker \widetilde{B}_{12} = \sum_j |e_j\rangle\langle e_j| \otimes \ker B_j$ and $E_N(\widetilde{A}_{12}) = \frac{1}{m} I \otimes A$, $E_N(\widetilde{B}_{12}) = \frac{1}{m} I \otimes B$. Then (c) implies $E_N(\widetilde{A}_{12})^{it} E_N(\widetilde{B}_{12})^{-it} P_{(\ker \widetilde{B}_{12})^\perp} = \widetilde{A}_{12}^{it} \widetilde{B}_{12}^{-it}$ for all t. Then by using Theorem 9 with $Q_1 = \widetilde{A}_{12}, Q_m = Q_2 = \widetilde{B}_{12}$, we can conclude that there are

positive matrices $A_0, B_0 \in M_d$ and $D_{12} \in (M_m \otimes M_d)^+$, such that $\ker D_{12} = \ker \widetilde{B}_{12}$, $[I \otimes A_0, D_{12}] = [I \otimes B_0, D_{12}] = 0$ and

$$\widetilde{A}_{12} = (I \otimes A_0)D_{12}, \qquad \widetilde{B}_{12} = (I \otimes B_0)D_{12} \tag{40}$$

Since $\widetilde{A}_{12}$, $\widetilde{B}_{12}$ are block diagonal, $D_{12} = \sum_j |e_j\rangle\langle e_j| \otimes D_j$ must also be block diagonal with $D_j \in M_d^+$, $\ker D_j = \ker B_j$, $[A_0, D_j] = [B_0, D_j] = 0$ for all j and

$$A_j = A_0 D_j, \qquad B_j = B_0 D_j. \tag{41}$$

Taking Tr_1 in (40) gives $A = A_0 D$ and $B = B_0 D$. Using this in (41) gives (38) which proves (d). The implications (d) $\Rightarrow$ (a) $\Rightarrow$ (b) are straightforward. **QED**

We return briefly to the case of arbitrary K. Note that if the condition (d) holds and $[D_j, K] = 0$ for all j, then $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ for all $p \in (0, 2)$, this gives a sufficient, but not necessary, condition for equality if $K \neq I$. The next result reduces the case of K unitary to $K = I$. Then, we can apply the conditions of Theorem 10 to A_j and $K B_j K^*$.

Theorem 11 *If K is unitary, then $J_p(K, A, B) = \sum_j J_p(K, A_j, B_j)$ if and only if $J_p(I, A, KBK^*) = \sum_j J_p(I, A_j, K B_j K^*)$*

Proof: When K is unitary, then $KB^pK^* = (KBK^*)^p$ which implies $J_p(K, A, B) = J_p(I, A, KBK^*)$. **QED**

One can try to extend the results of this section to the case $\|K\| \leq 1$, and hence to all K, by using the unitary dilation

$$\mathcal{U} = \begin{pmatrix} K & L \\ -L & K \end{pmatrix}$$

where $L = U(1 - |K|^2)^{1/2}$ and $K = U|K|$ is the polar decomposition. Then, with

$$\mathcal{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$$

we have $J_p(K, A, B) = J_p(\mathcal{U}, \mathcal{A}, \mathcal{B})$, so that we may use Theorem 11 to get conditions for equality. But note that the conditions of Theorem 10 require that $\ker \mathcal{U}\mathcal{B}_j\mathcal{U}^* \subseteq \ker \mathcal{A}_j$ and it can be shown that this implies $P_{(\ker A_j)^\perp} K P_{(\ker B_j)^\perp} K^* = P_{(\ker A_j)^\perp}$, where $P_{\mathcal{N}}$ denotes a projection onto the subscripted space. In particular, if all A_j and B_j are invertible, this restricts us to unitary K.

3.4 Equality in monotonicity under partial trace

It is easy to see that when $A_{12} = A_1 \otimes A_2$ and $B_{12} = B_1 \otimes B_2$, then $J_p(I, A_{12}, B_{12}) = J_p(I, A_2, B_2)$ if and only if $A_1 = B_1$ with $\operatorname{Tr} A_1 = 1$. However, it is not necessary that $A_{12} = A_1 \otimes A_2$. The equality conditions are given by the following theorem.

Theorem 12 *Let $K_{12} = I_{12}$ and $A_{12}, B_{12} \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)^+$, with $\ker B_{12} \subseteq \ker A_{12}$. Equality holds in* (28) *if and only if*

(i) $\mathcal{H}_2 = \bigoplus_n \mathcal{H}_n^L \otimes \mathcal{H}_n^R$,

(ii) $A_{12} = \bigoplus_n A_n^L \otimes A_n^R$ *with* $A_n^L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)^+$ *and* $A_n^R \in \mathcal{B}(\mathcal{H}_n^R)^+$,

(iii) $B_{12} = \bigoplus_n B_n^L \otimes B_n^R$ *with* $B_n^L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)^+$ *and* $B_n^R \in \mathcal{B}(\mathcal{H}_n^R)^+$,

(iv) $A_n^L = B_n^L$ *for all* n.

Proof: Let us denote $A_j = \frac{1}{d_1}\mathcal{W}_j A_{12} \mathcal{W}_j^*$, $B_j = \frac{1}{d_1}\mathcal{W}_j B_{12} \mathcal{W}_j^*$, with $\mathcal{W}_j$ defined as in the proof of Theorem 7. Then we get that equality in (28) is equivalent to

$$J_p(I_{12}, \sum_j A_j, \sum_j B_j) = \sum_j J_p(I_{12}, A_j, B_j)$$

By Theorem 10, equality for some p implies equality for all p, so that $J_p(I_{12}, A_{12}, B_{12}) = J_p(I_2, \operatorname{Tr}_1 A, \operatorname{Tr}_1 B) = J_p(I_{12}, E_N(A_{12}), E_N(B_{12}))$ for $p \in (0,1)$, where N is the subalgebra $I_1 \otimes \mathcal{B}(\mathcal{H}_2) \subseteq \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. Hence N is sufficient for $\{A_{12}, B_{12}\}$ and, by Theorem 9, there are some $A_R, B_R \in \mathcal{B}(\mathcal{H}_2)^+$ and $D \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)^+$, $\ker D = \ker B_{12}$, such that $[(I_1 \otimes A_R), D] = [(I_1 \otimes B_R), D] = 0$ and

$$A_{12} = D(I_1 \otimes A_R), \quad B_{12} = D(I_1 \otimes B_R) \tag{42}$$

Now let M_1 be the subalgebra in $\mathcal{B}(\mathcal{H}_2)$, generated by A_R, B_R. Then $D \in (I_1 \otimes M_1)' = \mathcal{B}(\mathcal{H}_1) \otimes M_1'$ where M' denotes the commutant of M. There is a decomposition $\mathcal{H}_2 = \bigoplus_n \mathcal{H}_n^L \otimes \mathcal{H}_n^R$, such that

$$M_1' = \bigoplus_n \mathcal{B}(\mathcal{H}_n^L) \otimes 1_n^R, \qquad M_1 = \bigoplus_n 1_n^L \otimes \mathcal{B}(\mathcal{H}_n^R)$$

and $D = \bigoplus_n D_n \otimes 1_n^R$, where $D_n \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)$. Since $A_R, B_R \in M_1$, we get the result, with $A_n^L = B_n^L = D_n$. The converse can be verified directly. **QED**

Applying this result in the case $A_{12} \mapsto A_{123}$ and $B_{12} \mapsto A_{12} \otimes I_3$ gives equality conditions in (34). Since these are independent of p, they are identical to the conditions, first given in [13], for equality in SSA (35) which corresponds to $p = 1$.

Corollary 13 *Equality holds in* (34) *if and only if*

(i) $\mathcal{H}_2 = \bigoplus_n \mathcal{H}_n^L \otimes \mathcal{H}_n^R$.

(ii) $A_{123} = \bigoplus_n A_n^L \otimes A_n^R$ *with* $A_n^L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_n^L)$ *and* $A_n^R \in \mathcal{B}(\mathcal{H}_n^R \otimes \mathcal{H}_3)$

Proof: It suffices to let $A_{12} \to A_{123}$ and $B_{12} \to A_{12} \otimes I_3$ in Theorem 12. **QED**

To apply these results in Section 4, it is useful to observe that condition (ii) in Corollary 13 above can be written as

$$A_{123} = (F_L \otimes I_3)(I_1 \otimes F_R) \tag{43}$$

with $F_L \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)^+$, $F_R \in \mathcal{B}(\mathcal{H}_2 \otimes \mathcal{H}_3)^+$, $[F_L \otimes I_3, I_1 \otimes F_R] = 0$. Combining this with part (d) of Theorem 10 gives the following useful result, which essentially allows us to bypass the need to apply Theorem 10 to $J_p(I, A_j, \mathcal{W}_n A_j \mathcal{W}_n)$.

Corollary 14 *Let $A_j \in M_{d_1} \otimes M_{d_2}$, $A = \sum A_j$. Then*

$$J_p(I_{12}, A, (\mathrm{Tr}_2 A) \otimes I_2) = \sum_j J_p(I_{12}, A_j, (\mathrm{Tr}_2 A_j) \otimes I_2) \tag{44}$$

if and only if there are $D_j \in M_{d_1}^+$, such that $\ker D_j = \ker \mathrm{Tr}_2 A_j$, *$[A_j, D_j \otimes I] = 0$ and $A_j = A(D^{-1} D_j \otimes I)$ with $D = \sum_j D_j$.*

Proof: Let $\widetilde{A}_{123} = \sum_j |e_j\rangle\langle e_j| \otimes A_j \in M_m \otimes M_{d_1} \otimes M_{d_2}$, then $A = \widetilde{A}_{23} \in M_{d_1} \otimes M_{d_2}$ and (44) can be written as

$$J_p(I_{23}, \widetilde{A}_{23}, \widetilde{A}_2 \otimes I_3) = J_p(I_{123}, \widetilde{A}_{123}, \widetilde{A}_{12} \otimes I_3)$$

By (43), this is equivalent to the existence of F_L and F_R, $[(F_L \otimes I_3), (I_1 \otimes F_R)] = 0$, such that $\widetilde{A}_{123} = (F_L \otimes I_3)(I_1 \otimes F_R)$. Since $\widetilde{A}_{(1)(23)}$ is block-diagonal, F_L must be of the form $F_L = \sum_j |e_j\rangle\langle e_j| \otimes D_j$, so that $A_j = F_R(D_j \otimes I)$. Then $\mathrm{Tr}_2 A_j = D_j \mathrm{Tr}_2 F_R$ which implies that $\ker D_j \subseteq \ker \mathrm{Tr}_2 A_j$. If we let $P_j = P_{(\ker \mathrm{Tr}_2 A_j)^\perp}$, then P_j commutes with D_j and

$$A_j = (P_j \otimes I) A_j = (P_j D_j \otimes I) F_R,$$

so that we can assume that $\ker D_j = \ker \mathrm{Tr}_2 A_j$, by taking $P_j D_j$ instead of D_j. Taking Tr_1 of (43) gives $A = (D \otimes I_3) F_R = F_R (D \otimes I_3)$ so that $A_j = A(D^{-1} D_j \otimes I)$. **QED**

4 Equality in joint convexity of Carlen-Lieb

Carlen and Lieb [8] obtained several convexity inequalities from those of the map

$$\Upsilon_{p,q}(K, A) \equiv \mathrm{Tr}\,(K^* A^p K)^{q/p} \tag{45}$$

using an identity which we write only for $q = 1$ and $p > 1$ in our notation as

$$\Upsilon_{p,1}(K, A) = (p-1) \inf \left\{ J_p(K, A, X) + \tfrac{1}{p} \mathrm{Tr}\, X + \tfrac{1}{p(p-1)} \mathrm{Tr}\, K^* A K : X > 0 \right\} \tag{46}$$

We introduce the closely related quantity

$$\begin{aligned} \widehat{\Upsilon}_{p,1}(K,A) &= \inf\left\{J_p(K,A,X)+\tfrac{1}{p}\mathrm{Tr}\,X : X>0\right\} &(47)\\ &= \tfrac{1}{(p-1)}\left(\Upsilon_{p,1}(K,A)-\tfrac{1}{p}\mathrm{Tr}\,K^*AK\right) &(48)\end{aligned}$$

which is well-defined for all $p \in (0,2)$ and allows us to continue to treat the cases $p<1$ and $p>1$ simultaneously, as well as include the special case $p=1$ for which

$$\begin{aligned} \widehat{\Upsilon}_{1,1}(K,A) &= -\mathrm{Tr}\,K^*AK\log(K^*AK)+\mathrm{Tr}\,K^*(A\log A)K+\mathrm{Tr}\,K^*AK \\ &= S(K^*AK)+\mathrm{Tr}\,KK^*A\log A+\mathrm{Tr}\,K^*AK &(49)\end{aligned}$$

Since we are dealing with finite dimensional spaces, the infimum in (46) has a minimizer which satisfies

$$X_{\min} = (K^*A^pK)^{1/p}. \tag{50}$$

For fixed K, let X_j denote the minimizer associated with A_j. Then

$$\begin{aligned} \widehat{\Upsilon}_{p,1}(K,A_1)+\widehat{\Upsilon}_{p,1}(K,A_2) &= J_p(K,A_1,X_1)+\tfrac{1}{p}\mathrm{Tr}\,X_1+J_p(K,A_2,X_2)+\tfrac{1}{p}\mathrm{Tr}\,X_2 \\ &\geq J_p(K,A_1+A_2,X_1+X_2)+\tfrac{1}{p}\mathrm{Tr}\,(X_1+X_2) &(51)\\ &\geq \inf\left\{J_p(K,A_1+A_2,X)+\tfrac{1}{p}\mathrm{Tr}\,X : X>0\right\} \\ &= \widehat{\Upsilon}_{p,1}(K,A_1+A_2) &(52)\end{aligned}$$

which proves convexity of $\widehat{\Upsilon}_{p,1}$. Note that equality above requires both $X=\sum_j X_j$ and $J_p(K,A,X)=\sum_j J_p(K,A_j,X_j)$, where X is the minimizer associated with A.

Now we introduce some notation following the strategy in the published version of [8]. Let $|\mathbb{1}\rangle$ denote the vector $(1,1,\ldots,1)$ with all components 1 and $|e_1\rangle$ the vector $(1,0,\ldots,0)$. Define

$$\mathcal{K} = \tfrac{1}{d}I\otimes|\mathbb{1}\rangle\langle e_1| = \begin{pmatrix} I & 0 & \ldots & 0\\ I & 0 & \ldots & 0\\ \vdots & \vdots & & \vdots\\ I & 0 & \ldots & 0\end{pmatrix} \tag{53}$$

and

$$\mathcal{A}_j = \sum_k A_{jk}\otimes|e_k\rangle\langle e_k| = \bigoplus_k A_{jk} = \begin{pmatrix} A_{j1} & 0 & 0 & \ldots & 0\\ 0 & A_{j2} & 0 & \ldots & 0\\ 0 & 0 & A_{j3} & \ldots & 0\\ \vdots & & & \ddots & \vdots\end{pmatrix}, \tag{54}$$

and $\mathcal{A} = \sum_j \mathcal{A}_j = \sum_k A_k \otimes |e_k\rangle\langle e_k| = \oplus_k A_k$ with $A_k = \sum_j A_{jk}$. Then

$$\mathcal{K}^* \mathcal{A}^p \mathcal{K} = \big(\sum_k A_k^p \big) \otimes |e_1\rangle\langle e_1|$$

With this notation, we make some definitions following Carlen and Lieb but modified to allow a unified treatment of $p \in (0, 2)$.

$$\begin{aligned} \Phi_{(p,1)}(\mathcal{A}) = \Phi_{(p,1)}(A_1, A_2, A_3 \ldots) &\equiv \Upsilon_{p,1}(\mathcal{K}, \mathcal{A}) && (55) \\ &= \mathrm{Tr}\, \big(A_1^p + A_2^p + A_3^p + \ldots \big)^{1/p} \\ \widehat{\Phi}_{(p,1)}(\mathcal{A}) = \widehat{\Phi}_{(p,1)}(A_1, A_2, A_3 \ldots) &\equiv \widehat{\Upsilon}_{p,1}(\mathcal{K}, \mathcal{A}) && (56) \\ &= \tfrac{1}{(p-1)} \Big[\Phi_{(p,1)}(A_1, A_2, A_3 \ldots) - \tfrac{1}{p} \sum_k \mathrm{Tr}\, A_k \Big] \end{aligned}$$

The definitions of Φ and $\widehat{\Phi}$ apply only when $\mathcal{A}$ is a block diagonal matrix in $M_{d_1} \otimes M_{d_2}$. We now extend this to an arbitrary matrices $\mathcal{A}_{12} \in M_{d_1} \otimes M_{d_2}$.

$$\begin{aligned} \Psi_{(p,1)}(\mathcal{A}_{12}) &\equiv \mathrm{Tr}_1 \big(\mathrm{Tr}_2\, \mathcal{A}_{12}^p \big)^{1/p} && (57) \\ \widehat{\Psi}_{(p,1)}(\mathcal{A}_{12}) &\equiv \tfrac{1}{(p-1)} \big[\Psi_{(p,1)}(\mathcal{A}_{12}) - \tfrac{1}{p} \mathrm{Tr}\, \mathcal{A}_{12} \big] && (58) \end{aligned}$$

For $p = 1$, the formulas with hats are related to the conditional entropy, from which they differ by a constant

$$\begin{aligned} \widehat{\Phi}_{(1,1)}(A_1, A_2, A_3 \ldots) - \mathrm{Tr}\, \mathcal{A}_{12} &= -\mathrm{Tr}\, \big(\textstyle\sum_k A_k \big) \log \big(\textstyle\sum_k A_k \big) + \textstyle\sum_k A_k \log A_k \\ &= S\big(\textstyle\sum_k A_k \big) - S(\mathcal{A}_{12}) = J_1\big(I, \mathcal{A}_{12}, \mathrm{Tr}_2\, \mathcal{A}_{12} \otimes I_2 \big) && (59) \\ \widehat{\Psi}_{(1,1)}(\mathcal{A}_{12}) - \mathrm{Tr}\, \mathcal{A}_{12} &= S(\mathcal{A}_1) - S(\mathcal{A}_{12}) = H(\mathcal{A}_{12}, \mathcal{A}_1 \otimes I_2) && (60) \end{aligned}$$

When $\mathcal{A}_{12}$ is block diagonal, $\Psi_{(p,1)}(\mathcal{A}_{12}) = \Phi_{(p,1)}(\mathcal{A}_{12})$ with the understanding that $\mathrm{Tr}_2 \mathcal{A}_{12} = \sum_k A_k$. Now let W_n denote the generalized Pauli matrices as in Section 2.3, $\mathcal{W}_n = I_1 \otimes W_n$ and define

$$\mathcal{A}_{123} = \sum_n \mathcal{W}_n \mathcal{A}_{12} \mathcal{W}_n^* \otimes |e_n\rangle\langle e_n| = \bigoplus_n \mathcal{W}_n \mathcal{A}_{12} \mathcal{W}_n^* \qquad (61)$$

so that $\mathcal{A}_{123}$ is block diagonal with blocks $\mathcal{W}_n \mathcal{A}_{12} \mathcal{W}_n^*$. Then

$$d_2^{\frac{1+p}{p}} \Psi_{(p,1)}(\mathcal{A}_{12}) = \Phi(\mathcal{A}_{(12)(3)}) = \Phi\big(\mathcal{W}_1 \mathcal{A}_{12} \mathcal{W}_1^*,\, \mathcal{W}_2 \mathcal{A}_{12} \mathcal{W}_2^*,\, \ldots \big). \qquad (62)$$

It is straightforward to show that for $p \in (0, 2)$ the functions $\widehat{\Phi}_{(p,1)}(\mathcal{A})$ and $\widehat{\Psi}_{(p,1)}(\mathcal{A})$ are all convex in $\mathcal{A}$, inheriting this property from the quantities from which they are defined. In view of (59) and (60), the conditions for equality in the next two theorems are not surprising.

Theorem 15 *The function $\widehat{\Phi}_{(p,1)}(\mathcal{A})$ is convex in $\mathcal{A}$ for $p \in (0,2)$. Moreover, the following are equivalent:*

(i) $J_p\big(I, \mathcal{A}, (\mathrm{Tr}_2\, \mathcal{A}) \otimes I_2\big) = \sum_j J_p\big(I, \mathcal{A}_j, (\mathrm{Tr}_2\, \mathcal{A}_j) \otimes I_2\big)$,

(ii) There are matrices $D_j > 0$, $D = \sum_j D_j$, such that $[A_{jk}, D_j] = 0$, $\ker D_j = \ker(\sum_k A_{jk})$ and $A_{jk} = A_k D^{-1} D_j$,

(iii) $\widehat{\Phi}_{(p,1)}(A_1, A_2, A_3 \ldots) = \sum_j \widehat{\Phi}_{(p,1)}(A_{j1}, A_{j2}, A_{j3} \ldots)$.

Proof: It follows from Corollary 14 and the fact that $\mathcal{A}_j$ are block-diagonal that (i) $\Leftrightarrow$ (ii) and it is straightforward to verify that (ii) $\Rightarrow$ (iii). Moreover, (iii) implies (i) for $p = 1$, by (59). To show that (iii) implies (ii) for $p \neq 1$, observe that (iii), implies $\widehat{\Upsilon}_{p,1}(\mathcal{K}, \mathcal{A}) = \sum_j \widehat{\Upsilon}_{p,1}(\mathcal{K}, \mathcal{A}_j)$, and this implies

$$J_p(\mathcal{K}, \mathcal{A}, \mathcal{X}) = \sum_j J_p(\mathcal{K}, \mathcal{A}_j, \mathcal{X}_j) \tag{63}$$

where $\mathcal{X}_j = (\mathcal{K}^* \mathcal{A}_j^p \mathcal{K})^{1/p} = X_j \otimes |e_1\rangle\langle e_1|$ and $\sum_j \mathcal{X}_j = \mathcal{X} = (\mathcal{K}^* \mathcal{A}^p \mathcal{K})^{1/p} = X \otimes |e_1\rangle\langle e_1|$, with $X_j = (\sum_k A_{jk}^p)^{1/p}$ and $X = (\sum_k A_k^p)^{1/p}$. Since

$$\mathcal{K}^* \mathcal{A}_j^p \mathcal{K} \mathcal{X}_j^{1-p} = \sum_k A_{jk}^p X_j^{1-p} \otimes |e_1\rangle\langle e_1|,$$

with a similar expression for $\mathcal{K}^* \mathcal{A}^p \mathcal{K} \mathcal{X}^{1-p}$, we find

$$\sum_k J_p(I, A_k, X) = J_p(\mathcal{K}, \mathcal{A}, \mathcal{X}) = \sum_j J_p(\mathcal{K}, \mathcal{A}_j, \mathcal{X}_j) = \sum_{k,j} J_p(I, A_{jk}, X_j)$$

Convexity then implies that we must have

$$J_p(I, A_k, X) = \sum_j J_p(I, A_{jk}, X_j) \qquad \forall\ k. \tag{64}$$

Since $\ker X_j \subseteq \ker A_{jk}$, Theorem 10 implies that

$$A_k^{it} X^{-it} P_{(\ker X_j)^\perp} = A_{jk}^{it} X_j^{-it}, \quad \text{ for all } k, j, t \tag{65}$$

After writing $\widetilde{A}_k = \sum_j |e_j\rangle\langle e_j| \otimes A_{jk}$, $\widetilde{X} = \sum_j |e_j\rangle\langle e_j| \otimes X_j$, this reads

$$\widetilde{A}_k^{it} \widetilde{X}^{-it} = (\tfrac{1}{m} I \otimes \mathrm{Tr}_1 \widetilde{A}_k)^{it} (\tfrac{1}{m} I \otimes \mathrm{Tr}_1 \widetilde{X})^{-it} P_{(\ker \widetilde{X})^\perp},$$

so that, by Theorem 9, there are elements $B_k \in M_d^+$ and $D \in (M_m \otimes M_d)^+$, such that $\ker D = \ker \widetilde{X}$, $[(I \otimes B_k), D] = 0$ and $\widetilde{A}_k = (I \otimes B_k) D$. As before, one finds $D = \sum_j |e_j\rangle\langle e_j| \otimes D_j$ for some $D_j \in M_d^+$ which implies (ii). **QED**

Theorem 16 *The function $\widehat{\Psi}_{(p,1)}(\mathcal{A}_{12})$ is convex in $\mathcal{A}_{12}$ for $p \in (0,2)$. Moreover, if we let $\mathcal{A}_{123}$ denote the block diagonal matrix with blocks $\mathcal{W}_n \mathcal{A} \mathcal{W}_n^*$, the following are equivalent:*

(i) $J_p(I, \mathcal{A}_{123}, \mathcal{A}_1 \otimes I_{23}) = \sum_j J_p\big(I, (\mathcal{A}_{123})_j, (\mathcal{A}_1)_j \otimes I_{23}\big)$ with $\mathcal{A}_{123}$ defined by (61),

(ii) There are matrices $D_j \in M_{d_1}^+$, $D = \sum_j D_j$, such that $\ker D_j = \ker(\mathcal{A}_1)_j$, $[\mathcal{A}_j, D_j \otimes I] = 0$ and $\mathcal{A}_j = \mathcal{A}(D^{-1} D_j \otimes I)$.

(iii) $\widehat{\Psi}_{(p,1)}(\mathcal{A}) = \sum_j \widehat{\Psi}_{(p,1)}(\mathcal{A}_j)$.

Proof: It follows from the definition of $\mathcal{A}_{123}$, that $d_2^{\frac{1+p}{p}} \Psi_{(p,1)}(\mathcal{A}) = \Phi(\mathcal{A}_{123})$. The equivalence (i) $\Leftrightarrow$ (iii) follows immediately from Theorem 15, and (i) $\Leftrightarrow$ (ii) can be shown to follow from Corollary 14. **QED**

Theorem 17 *The following monotonicity inequalities hold,*

$$\widehat{\Psi}_{(p,1)}(\mathcal{A}_{23}) \leq \widehat{\Psi}_{(p,1)}(\mathcal{A}_{123}), \quad p \in (0,2) \tag{66a}$$

$$\Psi_{(p,1)}(\mathcal{A}_{23}) \geq \Psi_{(p,1)}(\mathcal{A}_{123}), \quad p \in (0,1) \tag{66b}$$

$$\Psi_{(p,1)}(\mathcal{A}_{23}) \leq \Psi_{(p,1)}(\mathcal{A}_{123}), \quad p \in [1,2) \tag{66c}$$

Moreover, equality holds if and only if the conditions of Corollary 13 are satisfied.

Proof: It suffices to give the proof for $\widehat{\Psi}$ since the other inequalities follow immediately. The argument is similar to that for Theorem 7. Let W_n denote the generalized Pauli matrices of Section 2.3, but now let $\mathcal{W}_n = W_n \otimes I_{23}$. Then the convexity of $\widehat{\Psi}_{(p,1)}(\mathcal{A}_{23})$ implies

$$\begin{aligned}
\widehat{\Psi}_{(p,1)}(\mathcal{A}_{23}) &= \tfrac{1}{d_1} \widehat{\Psi}_{(p,1)}(I_1 \otimes \mathcal{A}_{23}) \\
&= \tfrac{1}{d_1} \widehat{\Psi}_{(p,1)}\big(\tfrac{1}{d_1} \textstyle\sum_n \mathcal{W}_n \mathcal{A}_{123} \mathcal{W}_n\big) \\
&\leq \tfrac{1}{d_1^2} \sum_n \widehat{\Psi}_{(p,1)}(\mathcal{W}_n \mathcal{A}_{123} \mathcal{W}_n) \ = \ \widehat{\Psi}_{(p,1)}(\mathcal{A}_{123})
\end{aligned}$$

where we used the invariance of $\widehat{\Psi}$ under unitaries of the form $U_1 \otimes I_{23}$. In the case $p = 1$, it follows from (60) that $\widehat{\Psi}_{(1,1)}(\mathcal{A}_{23}) \leq \widehat{\Psi}(1,1)(\mathcal{A}_{123})$ becomes

$$S(\mathcal{A}_2) - S(\mathcal{A}_{23}) \leq S(\mathcal{A}_{12}) - S(\mathcal{A}_{123}) \tag{67}$$

which is SSA. Because the equality conditions in Theorem 16 are independent of p, they are identical to those for SSA, which are given in Corollary 13. **QED**

The Carlen-Lieb triple Minkowski inequality for the case $q = 1$ is an immediate corollary of Theorem 17. Observe that

$$\begin{aligned} \mathrm{Tr}_3\,\mathrm{Tr}_1\big(\mathrm{Tr}_2\mathcal{A}_{123}^p\big)^{1/p} &= \Psi_{(p,1)}(\mathcal{A}_{(13),(2)}) \qquad (68a) \\ \mathrm{Tr}_3\left[\mathrm{Tr}_2(\mathrm{Tr}_1\,\mathcal{A}_{123})^p\right]^{1/p} &= \Psi_{(p,1)}(\mathcal{A}_{32}) \qquad (68b) \end{aligned}$$

so that it follows immediately from (66c) that

$$\mathrm{Tr}_3\left[\mathrm{Tr}_2(\mathrm{Tr}_1\,\mathcal{A}_{123})^p\right]^{1/p} = \Psi_{(p,1)}(\mathcal{A}_{32}) \le \Psi_{(p,1)}(\mathcal{A}_{132}) = \mathrm{Tr}_3\,\mathrm{Tr}_1\big(\mathrm{Tr}_2\mathcal{A}_{123}^p\big)^{1/p} \qquad (69)$$

for $1 < p \le 2$ and from (66b) that the inequality reverses for $0 < p < 1$. Moreover, the conditions for equality are again independent of p and identical to those for equality in SSA, given in Corollary 13.

5 Final remarks

It should be clear that the results in Section 2 are not restricted to $J_p(K, A, B)$. The function $g_p(x)$ given in (6) can be replaced by any operator convex function of the form $g(x) = xf(x)$ with f operator monotone on $(0, \infty)$. Moreover, if the measure $\nu(t)$ in (17) is supported on $(0, \infty)$, then the conditions for equality are identical to those in Section 3.

In particular, our results go through with g_p replaced by $\widetilde{g}_p$ and $J_p(I, A, B)$ replaced by $\widetilde{J}_p(I, A, B)$, which is well-defined for $p \in [-1, 1)$ with $\widetilde{J}_0(I, A, B) = H(B, A)$. Thus our results can be extended to all $p \in (-1, 2)$. The case $p = 2$ reduces to the convexity of $(A, X) \mapsto \mathrm{Tr}\, X^* A^{-1} X$ with $A > 0$ proved in [24]. One can show that equality holds if and only if $X_j = A_j T \quad \forall\, j$ with $T = A^{-1}X$. We recently learned that Kiefer [16] proved the $p = 2$ convexity, by a different method, much earlier and also found these equality conditions.

There have been various attempts, e.g., the Renyi [35] and Tsallis [40] entropies, to generalize quantum entropy in a way that gives the usual von Neumann entropy at $p = 1$. In this paper we have considered two extensions of the conditional entropy involving an exponent $p \in (0, 2)$, namely,

- $J_p(I, A_{12}, A_1)$ which gives $\mathrm{Tr}\, A_{23}^p A_2^{1-p} \begin{array}{c}\le\\ \ge\end{array} \mathrm{Tr}\, A_{123}^p A_{12}^{1-p} \quad \begin{array}{c} p \in (0,1) \\ p \in (1,2) \end{array}$ and can be thought of as a pseudo-metric; and
- $\widehat{\Psi}_{(p,1)}(\mathcal{A}_{12})$ which gives $\mathrm{Tr}_2(\mathrm{Tr}_3\mathcal{A}_{23}^p)^{1/p} \begin{array}{c}\ge\\ \le\end{array} \mathrm{Tr}_{12}(\mathrm{Tr}_3\mathcal{A}_{123}^p)^{1/p} \quad \begin{array}{c} p \in (0,1) \\ p \in (1,2) \end{array}$ and can be thought of as a pseudo-norm.

These expressions are quite different for $p \neq 1$, but arise from quantities with the same convexity and monotonicity properties, as well as the same equality conditions which are independent of p. Moreover, both yield SSA at $p = 1$ and the equality conditions for $p \neq 1$ are identical to those for SSA. This independence of non-trivial equality conditions on the precise form of the function seems remarkable.

If one uses $\widetilde{g}_p$ and $\widetilde{J}_p(I, A, B)$ from (10), then the inequalities above hold with $p \in (1, 2)$ replaced by $p \in (-1, 0)$ and SSA corresponds to $p = 0$.

A Proof of the key Schwarz inequality

For completeness, we include the proof of the joint convexity of $(A, B, X) \mapsto \operatorname{Tr} X^*(L_A + tR_B)^{-1}(X)$ when $A, B > 0$ and $t > 0$. Since this function is homogeneous of degree one, it suffices to prove subadditivity. Now let

$$M_j = (L_{A_j} + tR_{B_j})^{-1/2}(X_j) - (L_{A_j} + tR_{B_j})^{1/2}(\Lambda). \tag{70}$$

Then one can verify that

$$\begin{aligned} 0 &\leq \sum_j \operatorname{Tr} M_j^* M_j = \sum_j \langle M_j, M_j \rangle \\ &= \sum_j \operatorname{Tr} X_j^* (L_{A_j} + tR_{B_j})^{-1}(X_j) - \operatorname{Tr} \big(\textstyle\sum_j X_j^*\big)\Lambda \\ &\qquad\qquad - \operatorname{Tr} \Lambda^* \big(\textstyle\sum_j X_j\big) + \operatorname{Tr} \Lambda^* \textstyle\sum_j \big(L_{A_j} + tR_{B_j}\big)\Lambda. \end{aligned} \tag{71}$$

Next, observe that for any matrix W,

$$\sum_j \big(L_{A_j} + tR_{B_j}\big)(W) = \sum_j \big(A_j W + tW B_j\big) = L_{\sum_j A_j}(W) + tR_{\sum_j B_j}(W).$$

Therefore, inserting the choice $\Lambda = \big(L_{\sum_j A_j} + tR_{\sum_j B_j}\big)^{-1} \big(\sum_j X_j\big)$ in (71) yields

$$\operatorname{Tr} \big(\textstyle\sum_j X_j\big)^* \frac{1}{L_{\sum_j A_j} + tR_{\sum_j B_j}} \big(\textstyle\sum_j X_j\big) \leq \textstyle\sum_j \operatorname{Tr} X_j^* \frac{1}{L_{A_j} + tR_{B_j}} (X_j). \tag{72}$$

for any $t \geq 0$. **QED**

References

[1] N. I. Akheizer and I.M. Glazman, *Theory of Operators in Hilbert Space.* Vol. II (Frederik Ungar Publishing, NY, 1963).

[2] A. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs, **191** (American Mathematical Society and Oxford University Press, 2000).

[3] T. Ando, *Topics on Operator Inequalities* Sapporo Lecture Notes (1978).

[4] T. Ando, "Concavity of certain maps on positive definite matrices and applications to Hadamard products", *Lin, Alg. Appl.* **26**, 203-241 (1979)

[5] H. Araki, "Relative Entropy of States of von Neumann Algebras" *Publ RIMS Kyoto Univ.* **9**, 809–833 (1976).

[6] T. Bekjan, "On joint convexity of trace functions" *Lin, Alg. Appl.* **390**, 321–327 (2004).

[7] E. Carlen and E. Lieb, "A Minkowski type trace inequality and strong subadditivity of quantum entropy" *Amer. Math. Soc. Translations* **189**(2), 59–62 (1999). Reprinted in [21].

[8] Eric A. Carlen, Elliott H. Lieb "A Minkowski Type Trace Inequality and Strong Subadditivity of Quantum Entropy II: Convexity and Concavity" *Lett. Math. Phys.* **83**, 107–126 (2008). arXiv:0710.4167

[9] I. Devetak and J. Yard, "Exact Cost of Redistributing Multipartite Quantum States" *Phys. Rev. Lett.* **100**, 230501 (2008).

[10] W. F. Donoghue Jr., *Monotone Matrix Functions and Analytic Continuation* (Springer, 1974).

[11] E. G. Effros, "A Matrix Convexity Approach to Some Celebrated Quantum Inequalities" *Proc. Natl. Acad. Sci.* **106**, 1006–1008 (2009). arXiv:0802.1234

[12] H. Epstein, "Remarks on Two Theorems of E. Lieb" *Commun. Math. Phys.* **31**, 317–325 (1973).

[13] P. Hayden, R. Jozsa, D. Petz, A. Winter "Structure of states which satisfy strong subadditivity of quantum entropy with equality" *Commun. Math. Phys.* **246**, 359–374, (2004)4 arXiv:quant-ph/0304007

[14] A. Jenčová and D. Petz "Sufficiency in quantum statistical inference" *Commun. Math. Phys.* **263**, 259–276 (2006). arXiv:math-ph/0412093

[15] A. Jenčová and D. Petz “Sufficiency in quantum statistical inference. A survey with examples” *J. Infin. Dimens. Anal.and Quantum Prob. Relat. Top.*, **9** 331–352 (2006). arXiv:quant-ph/0604091

[16] J. Kiefer, “Optimum experimental designs”, *J. Roy. Statist. Soc. Ser. B* **21** 272–310 (1959).

[17] O. Klein, “Zur quantenmechanischen begründung der zweiten hauptsatzes der wäremlehre” *Z. Phys.* **72**, 767–775 (1931).

[18] A. Lesniewski and M.B. Ruskai, “Relative Entropy and Monotone Riemannian Metrics on Non-Commutative Probability Space” *J. Math. Phys.* **40**, 5702–5724 (1999).

[19] S. Luo, N. Li and X. Cao “Relation between ‘no broadcasting’ for noncommuting states and ‘no local broadcasting’ for quantum correlations” *Phys. Rev. A* **79**, 054305 (2009).

[20] E. H. Lieb, “Convex trace functions and the Wigner-Yanase-Dyson conjecture”, *Adv. Math.* **11**, 267–288 (1973). Reprinted in [21].

[21] *Inequalities: Selecta of E. Lieb*, M. Loss and M.B. Ruskai, eds (Springer, 2002)

[22] E. H. Lieb and M.B. Ruskai, “ A Fundamental Property of the Quantum-Mechanical Entropy, *Phys. Rev. Lett.* **30**, 434–436 (1973). Reprinted in [21].

[23] E. H. Lieb and M.B. Ruskai, “Proof of the Strong Subadditivity of Quantum Mechanical Entropy” *J. Math. Phys.* **14**, 1938–1941 (1973). Reprinted in [21].

[24] E. H. Lieb and M.B. Ruskai “Some Operator Inequalities of the Schwarz Type” *Adv. Math.* **12**, 269–273 (1974). Reprinted in [21].

[25] E. H. Lieb and W. Thirring, “Inequalities for the Moments of the Eigenvalues of the Schrödinger Hamiltonian and Their Relation to Sobolev Inequalities”, in *Studies in Mathematical Physics*, E. Lieb, B. Simon, A. Wightman eds., pp. 269–303 (Princeton University Press, 1976). Reprinted in [21].

[26] G. Lindblad, “Expectations and Entropy Inequalities” *Commun. Math. Phys.* **39**, 111–119 (1974).

[27] K. Löwner, “Über monotone Matrix Funktionen” *Math. Z.* **38**, 177–216 (1934).

[28] M. Mosonyi and D. Petz, Structure of sufficient quantum coarse-grainings, *Lett. Math. Phys.* **68**(2004), 19–30.

[29] H. Narnhofer and W. Thirring, "From Relative Entropy to Entropy" *Fizika* **17**, 257–265 (1986 ?).

[30] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993, 2nd edition 2004.

[31] D. Petz, "Quasi-Entropies for Finite Quantum Systems" *Rep. Math. Phys.* **23**, 57–65 (1986).

[32] D. Petz, "Sufficiency of channels over von Neumann algebras", *Quart. J. Math.* **39**, 907–1008 (1988), .

[33] D. Petz, "Sufficient Subalgebras and the Relative Entropy of States of a von Neumann Algebra" *Commun. Math. Phys.* **105**, 123–131 (1986).

[34] D. Petz, "Monotone Metrics on Matrix Spaces", *Lin. Alg. Appl.* **244**, 81–96 (1996).

[35] A. Rényi, "On measures of entropy and information" pp. 547–561 in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, Vol. I (Univ. California Press, Berkeley, 1961).

[36] M. B. Ruskai, "Inequalities for Quantum Entropy: A Review with Conditions for Equality" *J. Math. Phys.* **43**, 4358–4375 (2002); erratum **46**, 019901 (2005). (quant-ph/0205064).

[37] M. B. Ruskai "Another Short and Elementary Proof of Strong Subadditivity of Quantum Entropy" *Reports Math. Physics* **60**, 1–12 (2007). arXiv:quant-ph/0604206

[38] D. Ruelle, *Statistical Mechanics* (Benjamin, 1969).

[39] B. Simon, *The Statistical Mechanics of Lattice Gases* (Princeton Univ. Press, 1993).

[40] C. Tsallis, " Possible generalization of Boltzmann-Gibbs statistics" *J. Stat. Phys.* **52**, 47–4879 (1988).

[41] A. Wehrl "General Properties of Entropy" *Rev. Mod. Phys.* **50**, 221–260 (1978).

[42] E.P. Wigner and M. M. Yanase, "Information content of distributions" *Proc. Nat. Acad. Sci.* **49**, 910–918 (1963).

[43] E.P. Wigner and M. M. Yanase, "On the positive semi-definite nature of certain matrix expressions" *Canad. J. Math.* **16**, 397–406 (1964).